

Ирина Алдошина

Основы психоакустики

(подборка статей с сайта <http://www.625-net.ru>)

Содержание

Часть 1	3
Часть 2. Нелинейные свойства слуха	11
Часть 3 Слуховой анализ консонансов и диссонансов	14
Часть 4. Бинауральный слух и пространственная локализация	18
Часть 5. Бинуаральный слух (продолжение).....	25
Часть 6 Слуховая маскировка	28
Часть 7 Слуховая маскировка 2. Бинауральное маскирование	33
Часть 8. Слуховые пороги, часть 1	36
Часть 9. Слуховые пороги, часть 2.....	42
Часть 11. Аурализация - виртуальный звуковой мир.....	47
Часть 11. Громкость, ч. 1.....	53
Часть 12 Громкость сложных звуков, часть 2.....	58
Часть 13. Субъективные критерии оценки акустики помещений.ч. 1.....	62
Часть 13. Субъективные критерии оценки акустики помещений, часть 2.....	67
Часть 14. Тембр, часть 1.....	72
Часть 14 Тембр. Часть 2.....	78
Часть 14. Тембр, часть 3.....	85
Часть 15. Слуховое восприятие пространственных систем. Часть 1.....	92
Часть 15. Слуховое восприятие пространственных систем, часть 2.....	98
Часть 16. Взаимодействие акустических систем и помещения - стереофоническое прошлое и многоканальное будущее.....	104
Часть 17 Слух и речь. Часть 1.....	110
Часть 17. Слух и речь. Часть 2.....	119
Часть 17. Слух и речь Часть 3. Акустические характеристики речи.....	123
Часть 17. Слух и речь. Часть 4. Субъективные и объективные методы оценки разборчивости речи	128
Часть 17. Слух и речь Часть 4. Субъективные и объективные методы оценки разборчивости речи, часть 2.....	133
Часть 17. Слух и речь Часть 5 Акустические характеристики вокальной речи.....	142
Часть 17. Слух и речь, часть 5 Акустические характеристики вокальной речи (продолжение).....	149

Часть 1

Природа дала нам ноги и руки, чтобы спастись и защищаться - а мы изобрели спорт. Природа дала нам ощущение высоты, чтобы сортировать звуки окружающего мира - а мы изобрели музыку".

В. Хартман

Задача звукорежиссера - вместе с композитором и исполнителем - создать звуковой образ и передать его слушателю с помощью звукозаписи, звукоусиления, радиовещания, звукового сопровождения кино и телевидения и др..

Проблемами возникновения, передачи и восприятия звуков занимаются различные направления современной акустики, одним из которых является музыкальная акустика, которая изучает создание музыкальных звуков (акустика музыкальных инструментов, акустика речи и пения, электроакустика); передачу звуков (архитектурная акустика, звукозапись, усиление и вещание и др.) и восприятие звука (психоакустика - акустика слуха).

В конце 20 века именно психоакустика вышла на первый план. Научно-техническая революция открыла принципиально новые возможности работы со звуком, в том числе с помощью компьютерных музыкальных технологий. Она послужила базой для мощного развития аудиоиндустрии, создав новые средства передачи пространственной звуковой информации:

цифровое радиовещание, телевидение, звукозапись и т.д. В настоящее время достигнут принципиальный прогресс в том, как надо делать аппаратуру записи, передачи и воспроизведения звука. Однако конечным судьей этого процесса остается слуховая система, а принципы распознавания ею слухового образа еще до конца не изучены. Именно поэтому на эту науку сейчас обращены основное внимание и средства.

Основные задачи психоакустики - понять, как слуховая система расшифровывает звуковой образ, установить основные соответствия между физическими стимулами и слуховыми ощущениями, и выявить, какие именно параметры звукового сигнала являются наиболее значимыми для передачи семантической (смысловой) и эстетической (эмоциональной) информации.

Это принципиально важно как для дальнейшего развития аудиотехники, так и для музыкального искусства в целом (исполнительского творчества, совершенствования музыкальных инструментов, развития компьютерного музыкального синтеза и т.д.) и особенно для звукорежиссеров, поскольку понимание процессов формирования субъективного "слухового пространства" является необходимой базой их творчества.

1. Механизм работы слуховой системы

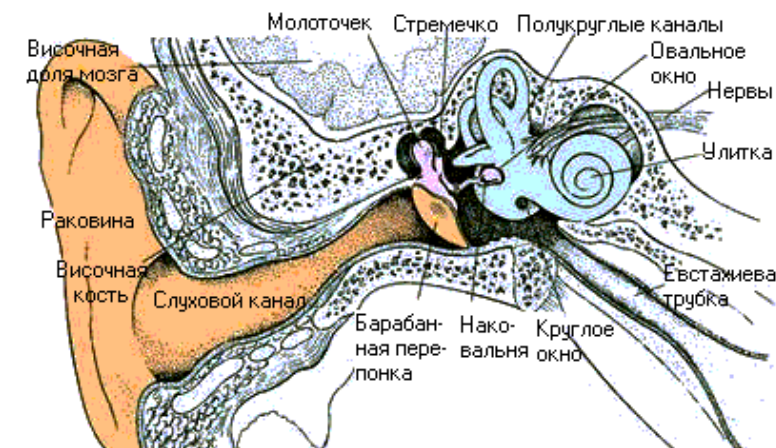


Рис 2 Структура периферической слуховой системы



Рис.1 Временная структура и спектральный состав ноты С4 (до первой октавы), сыгранной на скрипке

Звуковой сигнал любой природы может быть описан определенным набором физических характеристик: частота, интенсивность, длительность, временная структура, спектр и др. (Рис. 1). Им соответствуют определенные субъективные ощущения, возникающие при восприятии звуков слуховой системой: громкость, высота, тембр, биения, консонансы-диссонансы, маскировка, локализация-стереоэффект и т.п.

Слуховые ощущения связаны с физическими характеристиками неоднозначно и нелинейно, например, громкость зависит от интенсивности звука, от его частоты, от спектра и т.п.

Еще в прошлом веке был установлен закон Фехнера, подтвердивший, что эта связь нелинейна: "Ощущения пропорциональны отношению логарифмов стимула". Например, ощущения изменения громкости в первую очередь связаны с изменением логарифма интенсивности, высоты - с изменением логарифма частоты и т.д.

Всю звуковую информацию, которую человек получает из внешнего мира (она составляет примерно 25% от общей), он распознает с помощью слуховой системы и работы высших отделов мозга, переводит в мир своих ощущений, и принимает решения, как надо на нее реагировать.

Прежде чем приступить к изучению проблемы, как слуховая система воспринимает высоту тона, коротко остановимся на механизме работы слуховой системы. В этом направлении сейчас получено много новых и очень интересных результатов.

Слуховая система является своеобразным приемником информации и состоит из периферической части и высших отделов слуховой системы. Наиболее изучены процессы преобразования звуковых сигналов в периферической части слухового анализатора.

Периферическая часть

- это акустическая антенна, принимающая, локализирующая, фокусирующая и усиливающая звуковой сигнал;
- микрофон;
- частотный и временной анализатор;
- аналого-цифровой преобразователь, преобразующий аналоговый сигнал в двоичные нервные импульсы - электрические разряды.

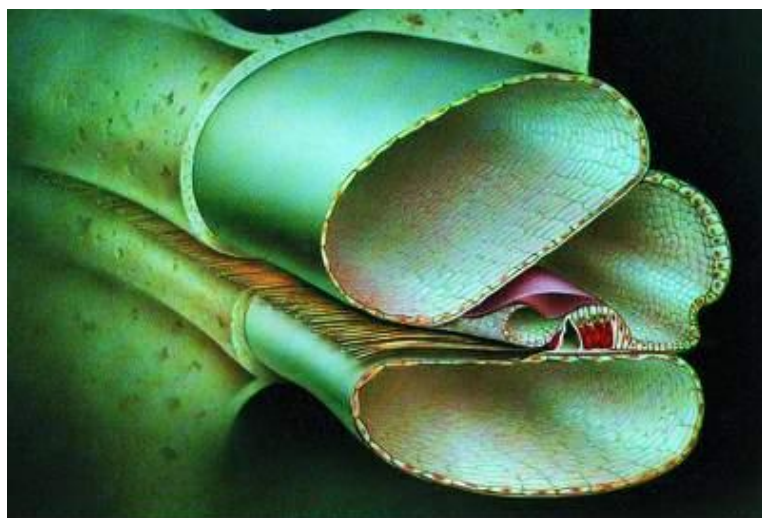
Общий вид периферической слуховой системы показан на рисунке 2. Обычно периферическую слуховую систему делят на три части: внешнее, среднее, и внутреннее ухо.

Внешнее ухо состоит из ушной раковины и слухового канала, заканчивающегося тонкой мембраной, называемой барабанной перепонкой. Внешние уши и голова - это компоненты внешней акустической антенны, которая соединяет (согласовывает) барабанную перепонку с внешним звуковым полем. Основные функции внешних ушей - бинауральное (пространственное) восприятие, локализация звукового источника и усиление звуковой энергии, особенно в области средних и высоких частот.

Слуховой канал представляет собой изогнутую цилиндрическую трубку длиной 22,5 мм, которая имеет первую резонансную частоту порядка 2,6 кГц, поэтому в этой области частот он существенно усиливает звуковой сигнал, и именно здесь находится область максимальной чувствительности слуха. Барабанная перепонка - тонкая пленка толщиной 74 мкм, имеет вид конуса, обращенного острием в сторону среднего уха. На низких частотах она движется как поршень, на более высоких - на ней образуется сложная система узловых линий, что также имеет значение для усиления звука.

Среднее ухо - заполненная воздухом полость, соединенная с носоглоткой евстахиевой трубой для выравнивания атмосферного давления. При изменении атмосферного давления воздух может входить или выходить из среднего уха, поэтому барабанная перепонка не реагирует на медленные изменения статического давления - спуск-подъем и т.п. В среднем ухе находятся три маленькие слуховые косточки: молоточек, наковальня и стремечко. Молоточек прикреплен к барабанной перепонке одним концом, вторым он соприкасается с наковальней, которая при помощи маленькой связки соединена со стремечком. Основание стремечка соединено с овальным окном во внутреннее ухо.

Среднее ухо выполняет следующие функции: согласование импеданса воздушной среды с жидкой средой улитки внутреннего уха; защита от громких звуков (акустический рефлекс); усиление (рычаговый механизм), за счет которого звуковое давление передаваемое во внутреннее ухо, усиливается почти на 38 дБ по сравнению с тем, которое попадает на барабанную



Улитка в разрезе

Рис 4 Бегущая волна на базилярной мембране

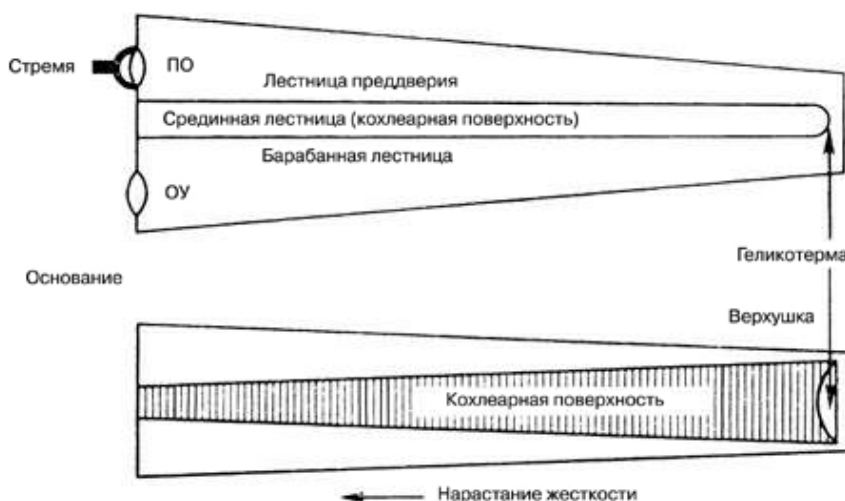
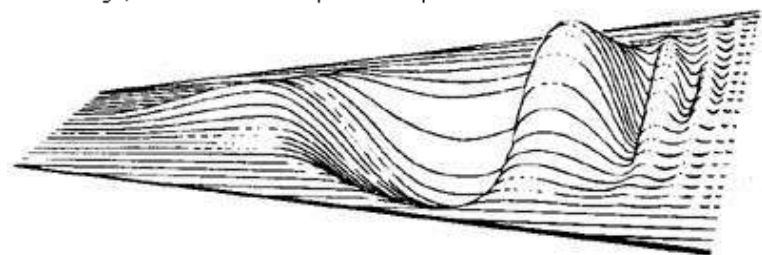


Рис 3 Схематическое изображение развернутой улитки (оп - окно преддверия; оу - окно улитки)

перепонку.

Внутреннее ухо находится в лабиринте каналов в височной кости, и включает в себя орган равновесия (вестибулярный аппарат) и улитку.

Улитка (cochlea) играет основную роль в слуховом восприятии. Она представляет собой трубку переменного сечения, свернутую три раза подобно хвосту змеи. В развернутом состоянии она имеет длину 3,5 см. Внутри улитка имеет чрезвычайно сложную структуру. По всей длине она разделена двумя мембранами на три полости: лестница преддверия, срединная полость и барабанная лестница (Рис. 3). Сверху срединная полость закрыта мембраной Рейсснера, снизу - базилярной мембраной. Все полости заполнены жидкостью. Верхняя и нижняя полости соединены через отверстие у вершины улитки (геликотрему). В верхней полости находится овальное окно, через которое стремечко передает колебания во внутреннее ухо, в нижней полости находится круглое окно, выходящее обратно в среднее ухо. Базилярная мембрана состоит из нескольких тысяч поперечных волокон: длина 32 мм, ширина у стремечка - 0,05 мм (этот конец узкий, легкий и жесткий), у геликотремы - ширина 0,5 мм (этот конец толще и мягче). На внутренней стороне базилярной мембраны находится орган Корти, а в нем - специализированные слуховые рецепторы - волосковые клетки. В поперечном направлении орган Корти состоит из одного ряда внутренних волосковых клеток и трех рядов наружных волосковых клеток. Между ними образуется тоннель. Волокна слухового нерва пересекают тоннель и контактируют с волосковыми клетками.

Слуховой нерв представляет собой перекрученный ствол, сердцевина которого состоит из волокон, отходящих от верхушки улитки, а наружные слои - от нижних ее участков. Войдя в ствол мозга, нейроны взаимодействуют с клетками различных уровней, поднимаясь к коре и перекрещиваясь по пути так, что слуховая информация от левого уха поступает в основном в правое полушарие, где происходит главным образом обработка эмоциональной информации, а от правого уха в левое полушарие, где в основном обрабатывается смысловая информация. В коре основные зоны слуха находятся в височной области, между обоими полушариями имеется постоянное взаимодействие.

Общий механизм передачи звука упрощенно может быть представлен следующим образом: звуковые волны проходят звуковой канал и возбуждают колебания барабанной перепонки. Эти колебания через систему косточек среднего уха передаются овальному окну, которое толкает жидкость в верхнем отделе улитки (лестнице преддверия), в ней возникает импульс давления, который заставляет жидкость переливаться из верхней половины в нижнюю через барабанную лестницу и геликотрему и оказывает давление на перепонку круглого окна, вызывая при этом его смещение в сторону, противоположную движению стремечка. Движение жидкости вызывает колебания базилярной мембраны (бегущая волна) (Рис. 4). Преобразование механических колебаний мембраны в дискретные электрические импульсы нервных волокон происходят в органе Корти. Когда базилярная мембрана вибрирует, реснички на волосковых клетках изгибаются, и это генерирует электрический потенциал, что вызывает поток электрических нервных импульсов, несущих всю необходимую информацию о поступившем звуковом сигнале в мозг для дальнейшей переработки и реагирования.

Высшие отделы слуховой системы (включая слуховые зоны коры), можно рассматривать как логический процессор, который выделяет (декодирует) полезные звуковые сигналы на фоне шумов, группирует их по определенным признакам, сравнивает с имеющимися в памяти образами, определяет их информационную ценность и принимает решение об ответных действиях.

2. Определение высоты звука

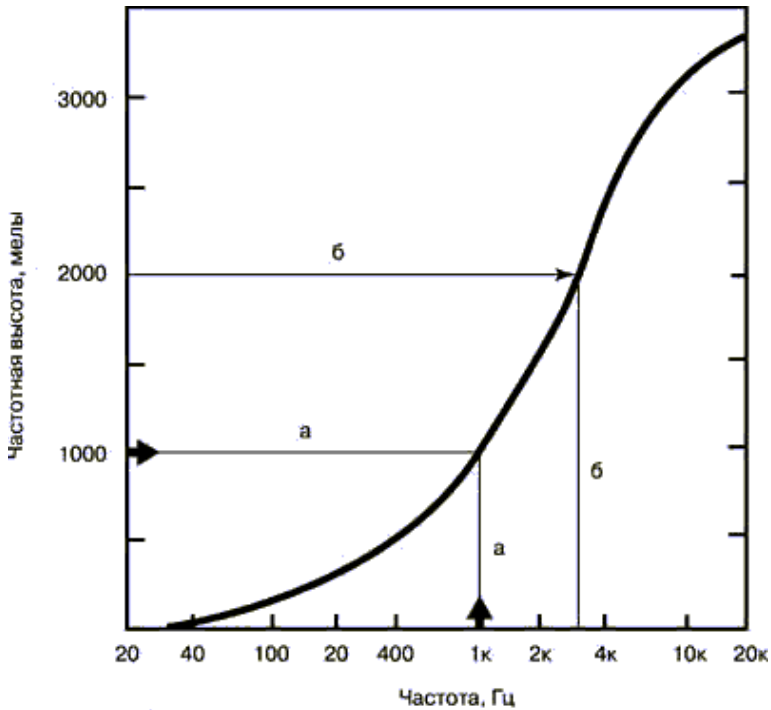


Рис 5 Зависимость высоты тона от частоты звука

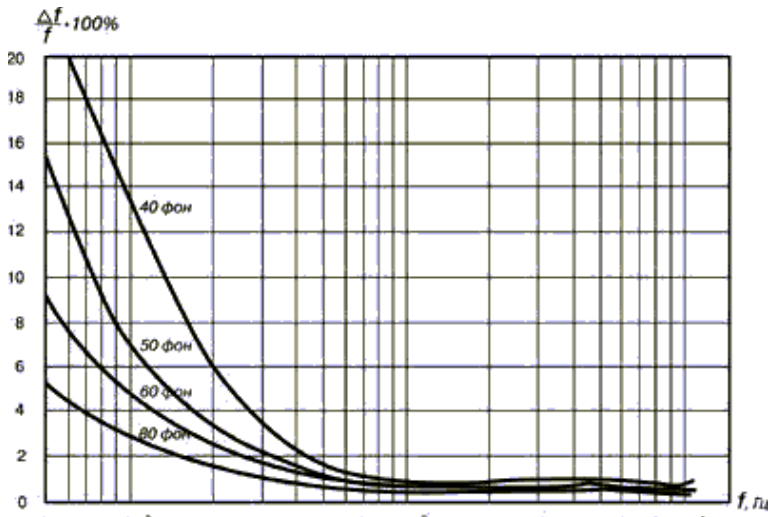


Рис 6 Кривые частотной разрешающей способности слуха

Nx fo(-1) (Гц)	2Гц	3Гц	4Гц	5Гц	6Гц	7Гц	8Гц	9Гц	10Гц
100	50	33,33	25	20	16,67	14,29	12,50	11,11	10
200	100	66,67	50	40	33,33	28,57	25	22,22	20
300	150	100	75	60	50	42,86	37,30	33,33	30
400	200	133,3	100	80	66,67	57,14	50	44,44	40
500	250	166,7	125	100	83,33	71,43	62,50	55,56	50
600	300	200	150	120	100	85,71	75	66,67	60
700	350	233,3	175	140	116,7	100	87,50	77,78	70
800	400	266,7	200	160	133,3	114,3	100	88,89	80
900	450	300	225	180	150	128,6	112,5	100	90
1000	500	333,3	250	200	166,7	142,9	125	111,1	100

Важнейшим свойством слуховой системы является возможность определения высоты звука. Это свойство имеет огромное значение для выделения и классификации звуков в окружающем звуковом пространстве, эта же способность слуховой системы лежит в основе восприятия интонационного аспекта музыки, то есть мелодии и гармонии.

В соответствии с международным стандартом ANSI- 1994 "Высота (Pitch) - это атрибут слухового ощущения в терминах, в которых звуки можно расположить по шкале от низких к высоким. Высота зависит главным образом от частоты звукового стимула, но она также зависит от звукового давления и от формы волны".

Таким образом, высота - это линейная классификация звуковых сигналов, в отличие от громкости, о которой можно сказать больше-меньше, т.е. это - относительная классификация.

Прежде всего, необходимо отметить, что слуховая система способна различать высоту звука только у периодических сигналов. Если это простое гармоническое колебание, например, синусоидальный сигнал от генератора, то период колебаний T определяет частоту $f = 1/T$, поэтому определяющим параметром для различения высоты является частота сигнала.

Если это сложный звук, то высоту слуховая система может присвоить по его основному тону, но только если он имеет периодическую структуру, т.е. спектр его состоит из гармоник (обертонов, частоты которых находятся в целочисленных отношениях). Если это условие не выполняется, то высоту тона определить слуховая система не может. Например, звуки таких инструментов как тарелки, гонги и др. не имеют определенной высоты.

Высота простых тонов

Изучение связи частоты звука и воспринимаемой высоты предпринималось еще Пифагором, а также многими известными физиками: Галилеем, Гельмгольцем, Омом и др. В настоящее время на основе тщательных экспериментов, в процессе которых слушателю предъявлялись два звука разной частоты с просьбой расположить их по высоте, установлена зависимость высоты тона от частоты сигнала, показанная на рисунке 5. Значения высоты отложено в специальных единицах - мелах. Один мел равен ощущаемой высоте звука частотой 1000 Гц при уровне 40 дБ (иногда для оценки высоты тона используется другая единица, барк = 100 мел). Как видно из рисунка, эта связь нелинейна - при увеличении частоты, например, в три раза (от 1000 до 3000 Гц), высота повышается только в два раза (от 1000 до 2000 мел). Нелинейность связи особенно выражена на низких и высоких частотах, в определенных пределах изменение высоты тона в мелах пропорционально логарифму частоты.

Многочисленные исследования были посвящены порогам различимости по высоте двух разных тонов, отличающихся по частоте. Результаты современных исследований представлены на рис.6, на котором видно, как слуховая система может различить по высоте два звука, отличающихся по частоте всего на 0,2%. Такая тонкая разрешающая способность слуха позволила установить, что ниже частоты 500 Гц можно выделить примерно 140 градаций высоты тона, в диапазоне от 500 Гц до 16 кГц - примерно 480 градаций высоты тона

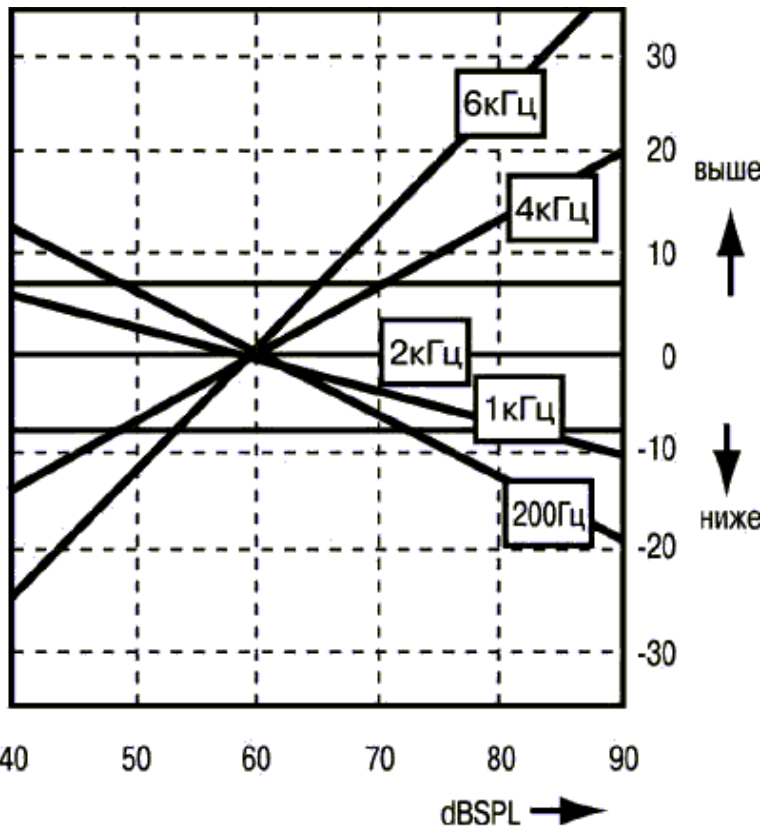


Рис 7 Зависимость высоты звука от его интенсивности

(всего 620 градаций). В европейской музыке инструменты с равномерно темперированной шкалой используют порядка 100 градаций высоты тонов. Но возможности слуховой системы гораздо больше - 620 градаций высоты, и это основа для развития современной микротоновой и спектральной музыки, то особенно продвинулось в связи с появлением компьютерных технологий.

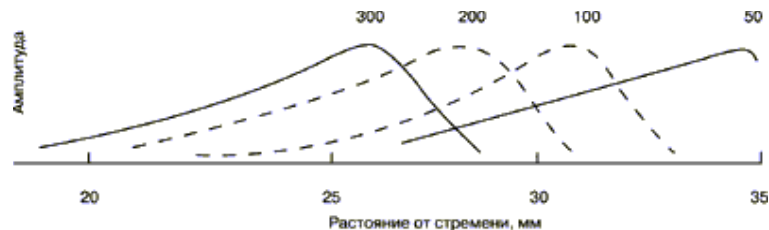


Рис 8 Зависимость высоты звука от его интенсивности

Ощущение высоты чистого тона (одной частоты) связано не только с частотой, но и с интенсивностью звука и его длительностью. Как показали различные исследования, при повышении интенсивности звука громкие низкие звуки кажутся еще ниже, а высокие звуки с повышением громкости кажутся слегка выше (зависимость показана на рисунке 7), для средних частот 1-2 кГц влияние интенсивности незаметно. Следует отметить, что эта зависимость незначительна, а для сложных музыкальных звуков почти незаметна. Это великое счастье для музыки, т.к. иначе при переходе от *pp* к *ff* звуковысотные отношения (мелодия и гармония) были бы нарушены.

Ощущение высоты тона зависит и от его длительности: короткие звуки воспринимаются как сухой щелчок, но при удлинении звука щелчок начинает давать ощущение высоты тона. Время, требуемое для перехода от щелчка к тону, зависит от частоты: для низких частот требуется для распознавания высоты тона примерно 60 мс, для частот от 1 до 2 кГц - 15 мс. Для сложных звуков это время увеличивается, для звуков речи оно может составлять 20-30 мс.

Высота сложных звуков

В музыке простые синусоидальные тоны практически не используются, каждый музыкальный тон имеет сложную структуру и состоит из основного тона и гармоник (пример ноты до на скрипке показан на рисунке 1).

Однако можно установить соответствие по высоте музыкального тона, например ноты ля первой октавы и чистого синусоидального сигнала с частотой 440 Гц. Высоты этих двух звуков будут одинаковыми, но тембры - разными. Это свидетельствует о том, что для сложных периодических сигналов высота присваивается по частоте основного тона - именно он имеет частоту 440 Гц.

В музыке используются другие шкалы для оценки высоты тона - музыкальные: полутоны, тоны, октавы и другие музыкальные интервалы. Следует отметить, что связь с психофизической шкалой высоты тона, построенной для чистых тонов, неоднозначна. До частоты примерно 5000 Гц увеличение высоты тона на октаву связано с удвоением частоты. Например, переход от ноты ля первой октавы к ноте ля второй октавы соответствует увеличению частоты от 440 до 880 Гц. Но выше частоты 5000 Гц это соответствие нарушается - чтобы получить ощущение увеличения высоты на октаву, надо увеличить соотношение частот почти в 10 раз, что следует иметь в виду при создании компьютерных композиций. Это дало основание некоторым ученым предложить две размерности высоты тона: психофизическую в мелах, пропорциональную в некоторых пределах логарифму частоты, установленную для чистых тонов (*pitch height*) и музыкальную, соответствующую названию нот (*pitch chroma*), которая может быть определена примерно до 5000 Гц. Следует отметить, что даже музыканты с абсолютным музыкальным слухом затрудняются в определении нот для звуков с частотой выше 5000 Гц. Это говорит о том, что механизмы восприятия высоты тона до 5000 Гц и выше - различны.

Для объяснения механизма восприятия высоты как простых, так и сложных звуков используются две теории: "теория места" и "временная теория".

3. Теория места

Теория места при восприятии высоты основана на способности базилярной мембраны выполнять частотный анализ сложного звука, т.е. действовать как спектральный анализатор. Базилярная мембрана организована тонотопически, т.е. каждый тон имеет свою топографию размещения. Как уже было указано выше, звуковой сигнал вызывает появление на мембране бегущей волны (Рис. 4), но специфика возбуждения состоит в том, что максимум смещения этой бегущей волны располагается в разных местах базилярной мембраны - низкие частоты имеют максимум смещения вблизи вершины мембраны, высокие - вблизи овального окна. Каждая частота имеет свое место максимума возбуждения на мембране (Рис. 8). В зависимости от спектрального состава на базилярной мембране возбуждаются различные участки. Возбуждаются волосковые клетки, находящиеся на этом месте, и их электрическая активность сообщает мозгу, какие частоты присутствуют в спектре. Таким образом, частота тона представлена в коде, основанном на том, нейроны каких участков активны, а каких - молчат. Физиологические исследования показывают, что тонотопическая организация

нейронов сохраняется во всех отделах мозга, вплоть до отделов слуховой коры. Логично допустить, что распознавание частоты и распознавание высоты есть результат тонотопического кодирования - в этом и заключается теория места.

При действии синусоидального сигнала в слуховом нерве формируется "образец возбуждения" - скорость разрядов нейронов как функция места на базилярной мембране. При этом пик этого образца движется вдоль мембраны при изменении частоты. Интересно отметить, что для того, чтобы слух различил два тона по высоте, необходимо, чтобы на базилярной мембране максимум смещения, соответствующий данным частотам, сместился всего на 52 мкм (если выразить в мелах, то одна градация высоты равна 3,9 мела).

Таким образом, можно считать, что периферическая слуховая система содержит банк полосовых фильтров ("слуховых фильтров") с перекрывающимися полосами (Рис. 8). Их ширина свыше 1кГц составляет примерно 10-17% от центральной частоты (например, на частоте 1000 Гц ширина полосы составляет 160 Гц). С шириной слуховых фильтров связано известное понятие "критической полосы" - внутри этой полосы звуковая информация интегрируется слухом; при выходе за пределы этой полосы происходит скачкообразное изменение слуховых ощущений, и это подтверждается экспериментами по маскировке, громкости, фазовой чувствительности и др.

При восприятии музыкального звука в соответствии с теорией места для слуховой системы существуют три возможности определения высоты:

Метод 1: локализовать место фундаментальной частоты и по нему определить высоту тона;

Метод 2: найти минимальную частотную разницу между соседними гармониками, которая равна фундаментальной частоте: $[(n+1)f_0] - (nf_0) = (nf_0) + (1f_0) - (nf_0) = f_0$, где $n = 1, 2, 3, \dots$ и принять ее за основу при распознавании высоты;

Метод 3: найти общий наибольший сомножитель, который получается при делении всех гармоник на последовательные целые числа, и использовать его как базу для определения частоты. Первой была предложена теория, по которой ощущаемая высота соответствует частоте только в том случае, если в звуковой волне присутствует энергия на этой частоте (второй закон Ома). Отсюда следовало, что присутствие фундаментальной частоты является обязательным для определения высоты звука. Первые сомнения в этой теории появились, когда стало возможным электрическим путем синтезировать спектры сложных звуков. В 1940 Шутен продемонстрировал, что ощущение высоты тона (сложной периодической волны) не изменится, если вырезать в музыкальном тоне фундаментальную частоту (Рис. 9).

Из этого следовало:

- присутствие фундаментальной частоты не обязательно для восприятия высоты;
- низшая частота не всегда является основой определения высоты.

Этот эксперимент получил название "феномен пропущенной фундаментальной" и доказал, что метод 1 не может служить единственной базой для определения высоты сложного тона, хотя он работает для большинства музыкальных, в том числе вокальных звуков.

Метод 2 дает возможность определить высоту тона по определению позиции соседних гармоник, даже если фундаментальная частота отсутствует. Для большинства музыкальных звуков соседние гармоники обычно присутствуют. Слуховая система, оценивая положение их максимумов на базилярной мембране, вычисляет частотную разницу между ними и по ней определяет высоту. Однако с помощью современных технических средств можно создать ситуацию, которую объяснить с помощью этого метода невозможно. Например, подаем звук, в котором присутствуют только нечетные гармоники $1f_0, 3f_0, 5f_0, 7f_0$, например, 100, 300, 500, 700 Гц и др. Если фундаментальная частота есть в спектре, то слух определяет высоту по ней $f_0 = 100$ Гц. Если ее вырезать, то расстояние между гармониками останется $2f_0$, но слух продолжает определять высоту тона, равную фундаментальной $f_0 = 100$ Гц.

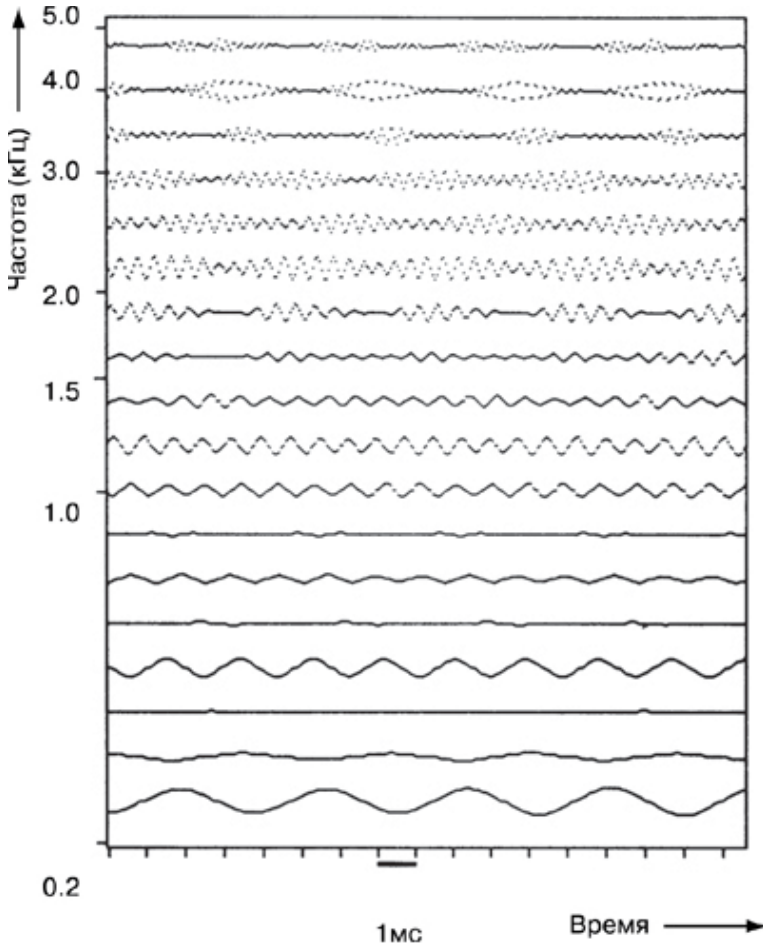


Рис 10а Вид звуковых волн для различных гармоник, выходящих из линейки слуховых фильтров (спектр ноты до-С4, сыгранной на скрипке)

Метод 3 позволяет объяснить и пропущенную фундаментальную и наличие только нечетных гармоник, т.к. от отсутствия каких-то гармоник общий наибольший сомножитель 100 Гц не меняется (см. таблицу). Этот метод позволяет также объяснить восприятие слабого ощущения высоты тона у колоколов и других источников квазипериодических тонов.

Механизм места разворачивает данную гармонику, если критическая полоса ее слухового фильтра, построенного на ней как на срединной частоте, достаточно узкая и соседние гармоники внутри этого фильтра не попадают. Если гармоники находятся настолько близко по частоте друг от друга, что внутри одного слухового фильтра попадает несколько гармоник, то они не разворачиваются. Какой бы ни была фундаментальная частота, слуховой механизм разворачивает только первые 6-7 гармоник - именно они и являются определяющими при определении высоты звука. Теория места создает базис для понимания того, как можно определить высоту путем анализа гармонического ряда, но эта теория не может объяснить ряд проблем, например, очень высокая точность определения высоты звука для тонов, чьи частотные компоненты не разворачиваются (т.е. звуки с гармониками выше седьмой).

4. Временная теория

Временная теория восприятия высоты базируется на анализе временной структуры звуковой волны (теория места на ее спектральном анализе). Эта теория использует синхронизацию разрядов нейронов органа Корти с фазой колебания базилярной мембраны (эффект запирающей фазы). При смещениях определенной точки мембраны в сторону расположения волосковых клеток в них возникает электрический потенциал, при смещении в противоположную сторону - потенциал отсутствует. Благодаря фазовому запирающему времени между импульсами в любом отдельном волокне будет равно целому числу 1, 2, 3... умноженному на период в основной звуковой волне. Нервные волокна кооперируются, чтобы кодировать частоты выше 300 Гц.

Основа временной теории - анализ формы волны в различных частях базилярной мембраны. Если рассматривать механизм частотного анализа на базилярной мембране как работу линейки фильтров различной ширины, то форма волны звукового сигнала, выходящего из этого набора фильтров, должна иметь вид, показанный на рисунке 10а. Например, если анализируется музыкальный тон с основной частотой 200 Гц, то выход из фильтра с центральной частотой 200 Гц имеет форму синусоидальной волны, т.к. эта гармоника разворачивается анализирующим фильтром. Аналогично разворачиваются этими фильтрами и все гармоники до пятой (около 1300 Гц). На выходе они имеют синусоидальную волну. Шестая гармоника (около 1560 Гц) имеет уже вариации амплитуды, но индивидуальные циклы еще видны. Волновая форма выходного сигнала для фильтра, центральная частота которого (в данном примере) выше шестой, не синусоидальная, т.к. гармоники не разворачиваются индивидуально, демонстрируя, что частотный диапазон полосового фильтра шире, чем расстояния между ними. По меньшей мере две гармоники комбинируются на выходе этого фильтра. Известно, что если две частоты находятся достаточно близко друг от друга, между ними возникают биения, т.е. одно колебание со средней частотой, равной разности частот. В данном случае, когда взаимодействуют две гармоники, этот период определяется фундаментальной частотой $T=1/f_0$. Таким образом, период всех волн, выходящих после фильтров с центральной частотой выше шестой гармоники и состоящих из соседних гармоник, будет одинаковым и равным $1/f_0$.

Минимальное время между импульсами от различных мест на базилярной мембране определяется периодом волны, выходящей от соответствующего фильтра. Для мест, которые соответствуют частотам от основной до шестой гармоники, минимальное время равно периоду данной гармоники. Для мест, соответствующих более высоким гармоникам, промежутки между импульсами равны периоду огибающей, т.е. основному тону (Рис. 10б). Таким образом, выше шестой гармоники разряды нейронов синхронизированы с формой огибающей, и период разрядов совпадает с периодом для фундаментальной частоты. Иными словами, для всех гармоник периоды разрядов или равны, или отличаются в целое число раз от частоты основного тона.

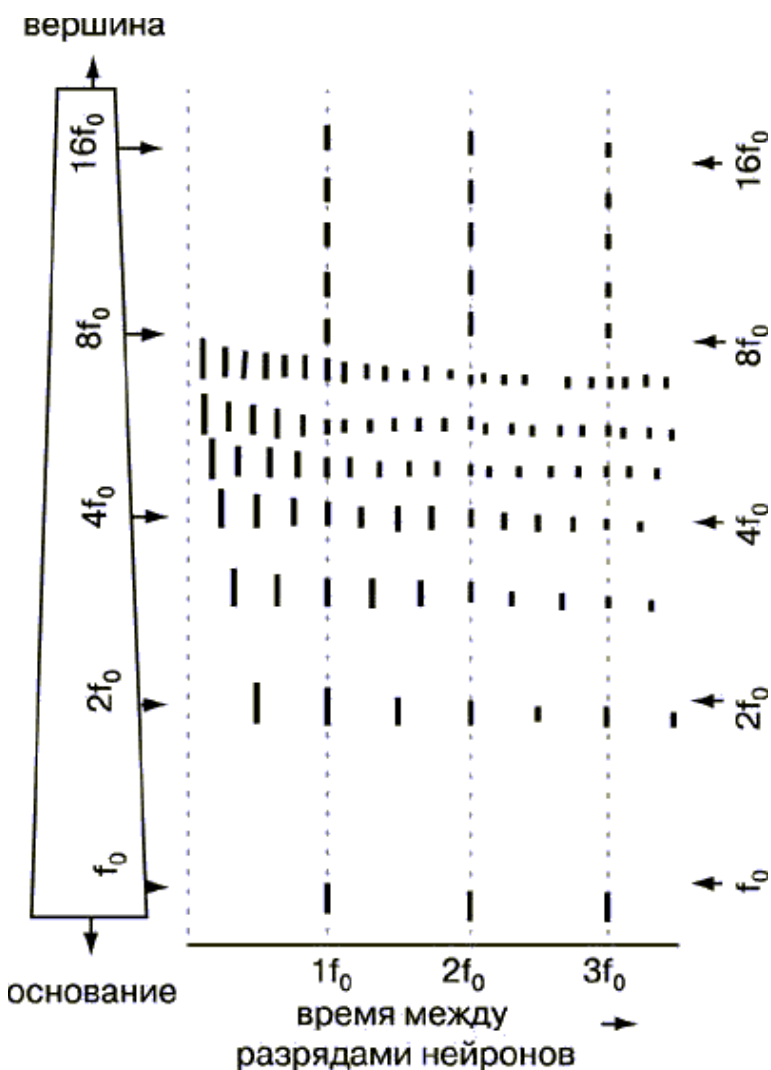


Рис 10б Периоды разрядов нейронов от разных мест базилярной мембраны

Это основа временной теории восприятия высоты тона: мозг определяет периодичность разрядов и по ним восстанавливает частоту основного тона.

Восприятие музыкальной высоты связано с оценкой временной формы звукового сигнала (за счет использования эффекта "фазового записания").

Временная теория позволяет понять, как найти фундаментальную частоту на основе анализа временных интервалов между нервными импульсами от различных мест на базилярной мембране и по ней определить высоту тона. Однако, временная теория не объясняет восприятия высоты тона на частотах выше 5000 Гц, т.к. эффект фазового записания не срабатывает на этих частотах. Вероятно, в этой области частот меняется механизм восприятия высоты тона.

Необходимо отметить, что на частотах выше 5 кГц в слуховой диапазон (до 20 кГц) попадают только две-три слышимых гармоники, этого слишком мало для слуха, поэтому, как уже было показано выше, восприятие высоты тона существенно обедняется и практически заканчивается восприятие музыкальной высоты (chroma pitch) тона (интонации). Вероятно, по этой причине, которая была интуитивно известна музыкантам, на большинстве музыкальных инструментов (рояль и др.) клавиатура заканчивается в области 5 кГц. На органе есть трубы, которые дают тон 8 кГц, но они употребляются только вместе с другими.

5. Современная теория восприятия высоты тона

Согласно современным теориям мозг принимает информацию от периферийной слуховой системы как за счет индикации места (частотный анализ), так и за счет информации о форме звуковой волны (временной анализ). Самостоятельно каждая теория, по-видимому, не может объяснить восприятие высоты полностью, т.к. та и другая информация передается по одним и тем же нервным волокнам.

Современная модель для восприятия высоты тона, объединяющая оба метода, показана на рисунке 11: сначала идет фильтрация сигнала по частоте с помощью развертки по месту, затем - анализ по межимпульсным интервалам (до шестой-седьмой гармоники они соответствуют периоду каждой гармоники), выше - по периоду огибающей. Поскольку период огибающей равен периоду основной частоты, то здесь различие высоты тона определяется только по месту возбуждения. Так определяется общий период, и по нему данному звуку присваивается определенная высота. Таким образом, обе теории дополняют друг друга.

Анализ восприятия высоты музыкального тона с помощью предложенной модели позволил получить ряд интересных результатов:

а) для музыкальных тонов с основной частотой от 100 до 400 Гц (с уровнем звукового давления не менее 50 дБ) основную роль в определении высоты тона играют первые пять-шесть гармоник (если их уровень превышает 10 дБ), т.е. те гармоники, которые разворачиваются слуховыми фильтрами;

б) звуковые сигналы, содержащие только очень высокие гармоники (свыше двадцатой), не вызывают ощущения высоты тона;

в) музыкальные сигналы, содержащие очень низкие частоты (с основной частотой ниже 50 Гц, например, звуки органа) вызывают ощущение высоты тона только по гармоникам, т.к. такие низкие частоты не вызывают смещений базилярной мембраны - они на ней не размещаются, им не хватает места. При этом наиболее существенную роль играют пятые-шестые гармоники;

г) основная частота звука, если она выше 1000 Гц, является доминантной компонентой в определении высоты тона;

д) музыкальные звуки, содержащие только неразвернутые гармоники (свыше шестой) могут дать ощущение высоты тона по огибающей, при этом слух дает достаточно тонкую дифференциацию сдвига максимума огибающей, т.е. точно чувствует высоту.

е) фазовые соотношения различных гармоник в музыкальном сигнале оказывают влияние на восприятие высоты, т.к. их изменение приводит к изменению структуры огибающей для высших неразвернутых гармоник. Для музыкальных сигналов, содержащих много низких и высоких гармоник, изменение фазовых соотношений

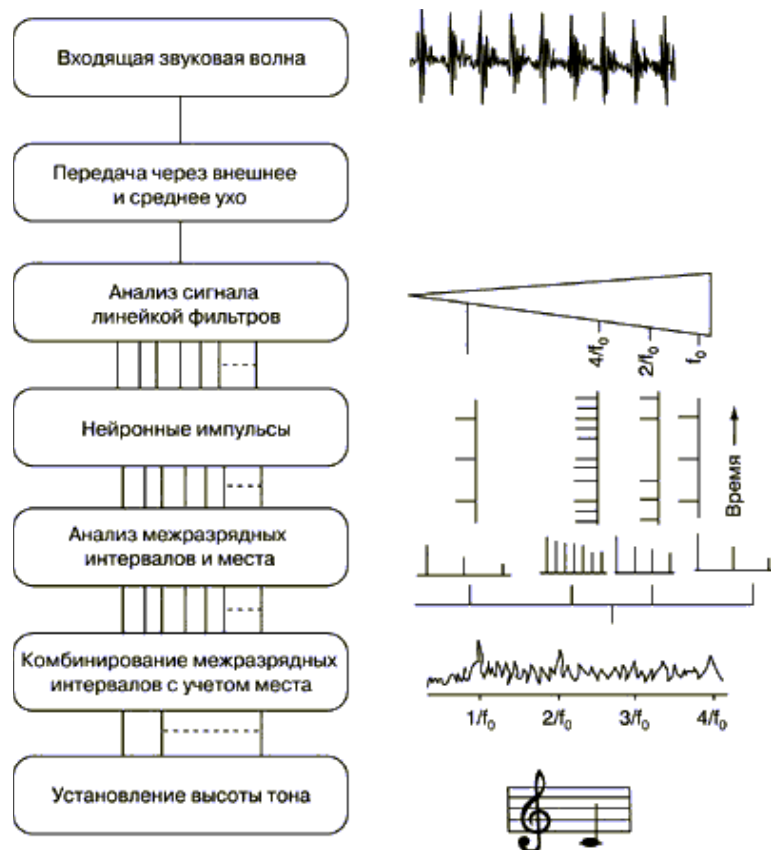


Рис 11 Модель восприятия высоты тона

может привести к улучшению четкости восприятия высоты, не вызывая ее сдвига (т.к. они не влияют на оценку низших развернутых гармоник). Для сигналов, содержащих в основном высокие гармоники, изменение их фазы может вызвать сдвиг высоты тона и изменение его четкости, т.к. может привести к сдвигу пиков в огибающей, по которым и определяется высота тона.

Таким образом, фазовые соотношения в музыкальном сигнале оказывают существенное влияние на звуковысотные отношения, что особенно важно учитывать в звукорежиссерской практике.

6. Высота тона и центральный процессор

Восприятие высоты тона для сложных музыкальных сигналов, как указано выше, начинается с анализа в периферической слуховой системе, где производится их частотный и временной анализ, а затем полученная информация передается в высшие отделы мозга - "центральный слуховой процессор", где полученная информация определенным образом группируется и осмысливается.

Мозг группирует несколько тонов (гармоник) с одинаковым частотным интервалом в одно ощущение высоты тона. Это принципиальное свойство слухового процессора (высших отделов коры головного мозга): из сложного внешнего звукового мира он выделяет звуки и группирует их по определенным признакам: по месту, по времени начала и конца, по периодичности повторений и т.п. Это связано с тем, что кратковременная память оперирует только шестью-семью символами и без группировки мозг не может принимать быстрых решений.

Современная психология утверждает, что мозг мыслит образами. По-видимому, музыкальные звуки также запоминаются в виде некоторых гармонических эталонов (шаблонов - template), которые формируются в детстве, аналогично звукам речи.

В настоящее время принята гипотеза, что центральный процессор, получив информацию от периферической слуховой системы о наличии компонент с кратными периодами в музыкальном звуке, группирует их и сравнивает с гармоническим шаблоном, в котором имеются все последовательные гармоники. Для каждого входного сигнала подбирается по фундаментальной частоте гармонический шаблон, который ему лучше подходит. В соответствии с этой моделью наиболее соответствующая фундаментальная частота выбранного шаблона и будет воспринимаемой высотой тона. Если два шаблона с разными фундаментальными частотами подходят к данному сигналу, можно ожидать услышать или неопределенную высоту или две высоты. В случае отсутствия фундаментальной частоты, сравнение производится по отдельным гармоникам. Если удастся подобрать хотя бы несколько гармоник, которые подходят под эталон, то по повторяющемуся интервалу между ними присваивается высота тона (виртуальная высота тона слышится, например, в звуке колоколов). Наиболее важными для синтеза ощущения высоты тона являются первые три - шесть развернутых гармоник. Компоненты сигнала, которые ведут себя аномально (например, одна гармоника включается-выключается или резко отличается от шаблона), выделяются центральным процессором и им присваивается отдельная высота.

Имеется много доказательств в поддержку данной гипотезы: например, при подаче разных гармоник в разные уши через телефоны (600 Гц в одно ухо и 800 Гц в другое), отчетливо слышен разностный тон высотой, соответствующей частоте 200 Гц, т.е. центральная система синтезирует высоту из гармоник в разных ушах. Другое доказательство, когда гармоники предъявляются одновременно: при последовательном включении третьей, четвертой и пятой гармоники по 40 мс с интервалом 10 мс, отчетливо слышался низкий тон с фундаментальной частотой и т.п.

Таким образом, в соответствии с этой моделью, гармоники собираются вместе, сравниваются центральным процессором с гармоническим эталоном (шаблоном) и по нему синтезируется высота музыкального тона.

Говоря о высоте комплексного тона, можно сказать, что "высота - великий консолидатор". Начиная с большого количества гармоник, процессор высоты объединяет их вместе в одно ощущение высоты. Слуховая организация определения высоты - основная часть осмысления звуков окружающего мира.

Важность определения высоты для слуховой системы не случайна и, вероятно, вовсе не результат стремления всего человечества сочинять музыку. Восприятие высоты играет центральную роль в определении индивидуальных объектов в акустическом мире и отделении их друг от друга. Окружающий мир наполнен конкурирующими звуками: интересными, угрожающими, шумовыми и др., все смешано вместе и слуховая система несет ответственность за их выделение и идентификацию. Высота есть главный идентификатор, позволяющий отделять данный звук от других объектов.

Часть 2. Нелинейные свойства слуха

Еще в 1714 году знаменитый скрипач Тартини заметил и описал странное явление: когда на скрипке громко проигрываются две ноты, иногда можно отчетливо слышать третий тон, которого не было у исполнителя. Такие же дополнительные тоны можно услышать на звуках флейты при двухголосном звучании. Это явление

вызвало большой интерес среди музыкантов и ученых, привело к постановке многочисленных экспериментов и позволило установить, что эти дополнительные "фантомные" тоны возникают непосредственно в слуховой системе и являются следствием ее нелинейности.

Интересно, что недопонимание этих процессов до сих пор приводит к недоразумениям, например, когда люди с тонким музыкальным слухом отчетливо слышат дополнительные тоны при исполнении некоторых аккордов, в то же время их коллеги могут их и не услышать. Особенно это касается людей старшего возраста, так как с возрастом слуховые пороги существенно меняются - чувствительность к высоким частотам уменьшается каждые десять лет примерно на 1000 Гц.

Учитывая огромные возможности для работы со звуком, которые предоставляют звукорежиссеру современные музыкальные технологии, им следует ознакомиться с теми звуковыми явлениями, к которым приводит нелинейность слуха.

По общему определению, система называется нелинейной, если выходной сигнал $Y(t)$ отличается от входного сигнала $X(t)$ наличием дополнительных спектральных составляющих. Обычно это имеет место, если связь между воздействующей силой (давлением) и откликом системы (смещением) является нелинейной. Практически вся электроакустическая аппаратура (громкоговорители, микрофоны, акустические системы и др.) является нелинейной (для оценки ее всегда нормируется коэффициент нелинейных искажений), однако эта нелинейность проявляется при достаточно больших уровнях входного сигнала.

Принципиальным отличием слухового аппарата является то, что он производит нелинейное преобразование входного звукового сигнала, как при большом его уровне, так и при очень малом, только механизмы этого преобразования различны.

Нелинейность слуха проявляется, прежде всего, в появлении "субъективных" или "слуховых" гармоник. При воздействии на барабанную перепонку достаточно громкого синусоидального звука с частотой f_0 в процессе его обработки в слуховом аппарате возникают гармоники этого звука с частотами $2f_0$, $3f_0$ и т.д. Например, если подать первичный тон с частотой 500 Гц, то можно услышать звуки с частотами 1000 Гц, 1500 Гц и т. д. Поскольку при объективных измерениях подводимого сигнала можно точно установить, что в спектре первичного воздействующего тона этих гармоник нет, они и получили название "субъективных" гармоник.

Наличие субъективных гармоник и их количественная оценка может быть выполнена с помощью прослушивания биений. Это явление возникает, если на систему подать два близких по частоте тона, например 1000 Гц и 1010 Гц; тогда вместо двух тонов будет отчетливо слышен один тон со средней частотой 1005 Гц, модулированный по амплитуде разностной частотой 10 Гц. Если разницу между двумя тонами увеличивать, то при разности частот выше 15 Гц биения исчезают; сначала начинают прослушиваться два тона с большой шероховатостью (как если бы звучали одновременно два ненастроенных музыкальных инструмента), затем отчетливо слышны два чистых тона. К биениям слух очень чувствителен, поэтому использование биений - основной метод настройки музыкальных инструментов.

Если к звуку, под действием которого возникают субъективные гармоники, например, 500 Гц, добавить второй скользящий тон, частоту и уровень которого можно плавно изменять, то при неточном совпадении частоты этого звука с частотой субъективной гармоники (например, 990 Гц и 1000 Гц) можно услышать на фоне громкого основного звука биения с разностной частотой ($f_{раз} = 10$ Гц), возникшие в результате взаимодействия скользящего звука и субъективной гармоники. Аналогичные измерения могут быть сделаны и для гармоник более высоких порядков. Наиболее резкие биения будут прослушиваться при равенстве их амплитуд. Поэтому, отрегулировав амплитуду давления скользящего звука до получения наиболее четких биений и измерив величину этого давления, можно определить величину субъективной гармоники. Эта техника называется "метод наилучших биений" - method of best beats. Полученные результаты позволили установить зависимость величины этих субъективных гармоник от уровня основного тона: например, при уровне тона с частотой 1000 Гц, равном 80 дБ SPL, уровень второй субъективной гармоники оказался равным 63 дБ. Уровень этих гармоник существенно зависит от уровня основного тона - только тогда, когда он становится ниже 40 дБ, эти гармоники становятся малыми, и возникает ощущение чистого тона.

При увеличении уровня интенсивности первичного тона величина субъективных гармоник резко возрастает. Это обстоятельство имеет существенное значение для восприятия слухом низкочастотных колебаний в диапазоне от 16 Гц до примерно 100 Гц.

Для того чтобы понять особенности слухового восприятия в этой области, вспомним, (см. предыдущую статью), что базилярная мембрана организована тонотопически, т. е. каждый тон имеет свою топографию размещения. В зависимости от спектрального состава на базилярной мембране возбуждаются различные участки, волосковые клетки находящиеся на этом месте возбуждаются и их электрическая активность сообщает мозгу, какие частоты присутствуют в спектре. Таким образом, базилярная мембрана выполняет функции спектрального анализатора с помощью линейки фильтров. Таким образом, звук с частотой 100 Гц воспринимается почти самым крайним участком базилярной мембраны близ ее верхушки, так что на базилярной мембране фактически нет участков, воспринимающих колебания более низких частот. Однако область слышимых звуков простирается значительно ниже (мы хорошо слышим частоты ниже 100 Гц). Предполагается, что звуки с частотой менее 100 Гц ощущаются не сами по себе, а из-за создаваемых ими

серий субъективных гармоник, попадающих в область частот выше 100 Гц, т. е. в конечном счете, из-за нелинейности слуха. Целый ряд фактов косвенно подтверждает это предположение, однако прямого подтверждения еще не найдено, так что пока это гипотеза.

Второй формой проявления нелинейности слуха является появление "субъективных комбинационных тонов". Как известно, если к нелинейной системе подвести два сигнала достаточно большого уровня с частотами f_1 и f_2 (например, 800 Гц и 1000 Гц), то нелинейные искажения вызовут появление комбинационных тонов с различными частотами, т. е. появляются вторичные комбинационные тоны: $f_2 - f_1$ и $f_2 + f_1$ (200 Гц и 1800 Гц), кубичные комбинационные тоны $2f_1 - f_2$ (600 Гц), $2f_2 - f_1$ (1200 Гц), $2f_1 + f_2$ (2600 Гц), $2f_2 + f_1$ (2800 Гц) и др. Для их количественной оценки также могут быть использованы "метод наилучших биений" или метод "погашений" (подается дополнительный сигнал с частотой комбинационного тона и подбирается его амплитуда и фаза, пока комбинационный тон не погасится, т.е. он подается в противофазе). Многочисленные эксперименты показали, что существуют особые комбинационные тоны, которые чаще всего прослушиваются при субъективных экспертизах: это разностные тоны с частотами $f_2 - f_1$ и $2f_1 - f_2$ (200 Гц и 600 Гц в нашем примере).

Простой разностный тон ведет себя, как в случае классической квадратичной нелинейности: он может быть услышан, если уровень первичных тонов больше, чем 50 дБ SPL; при равенстве уровней первичных тонов он увеличивается на 2 дБ; при возрастании уровня первичного тона на 1 дБ уровень этого тона не очень сильно зависит от отношения частот f_2 / f_1 .

В случае кубичного разностного тона установлено, что он возникает в основном при соотношении частот 11,3. В этом частотном диапазоне он может быть услышан при очень низком уровне первичных тонов: ниже 40 дБ SPL уровень f_2 может быть даже ниже 10 дБ. Увеличение амплитуды этого тона происходит не на 3 дБ при увеличении амплитуды первичного тона на 1 дБ (при равенстве их уровней) - как следовало бы при классической кубичной нелинейности - а существенно меньше. Все это заставляет предположить, что в образовании этих тонов участвуют некие дополнительные механизмы, которые мы рассмотрим далее.

Наконец, третий вид проявления нелинейности работы слухового аппарата - это нелинейная компрессия звукового сигнала. Уровень звукового сигнала в слышимом диапазоне меняется от 0 дБ до 120 дБ, т. е. амплитуда звукового давления меняется в 100 000 раз, в то же время динамический диапазон слухового нерва (от температурного шума до насыщения) составляет 1000. Поэтому, кроме функций спектрального анализатора, периферический слуховой аппарат выполняет функции нелинейного компрессора-усилителя.

Многочисленные исследования, особенно в последние годы, позволили получить ряд очень интересных результатов относительно механизмов возникновения нелинейности.

Как было показано в предыдущей статье, слуховой аппарат состоит из трех отделов - внешнего, среднего и внутреннего уха.

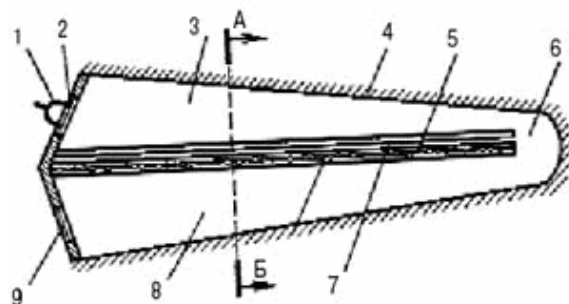
Экспериментально доказано, что преобразование сигнала во внешнем и среднем ухе - процесс линейный, основная причина нелинейности - в механизме работы внутреннего уха (улитки).

Улитка состоит из трех полостей, в которых находится жидкость (упрощенный разрез улитки показан на рис. 1). При ударе стремечка по мембране овального окна в жидкости возникает звуковой импульс, который распространяется из верхнего отдела в нижний и возбуждает базилярную мембрану. Исследования работы слуховой системы, выполненные знаменитым ученым Бекеши (Bekesy), за которые он получил Нобелевскую премию, показали, в частности, что при высоких уровнях сигнала в жидкости улитки образуются вихревые потоки. Поскольку ширина полостей разная, то этот процесс похож на образование околосонных завихрений, когда вода ударяется о берег (рис. 2а и рис. 2б). Появление этих завихрений искажает форму звукового импульса, а поскольку базилярная мембрана выполняет его спектральный анализ, то эти искажения и приводят к появлению дополнительных гармоник и комбинационных тонов.

Таким образом, первая причина возникновения нелинейных искажений - это гидродинамические процессы в жидкости улитки.

Рис. 1 Структура улитки (упрощенная)

1 - стремечко, 2 - овальное окно, 3 - вестибулярный ход, 4 - костная оболочка, 5 - орган Корти, 6 - отверстие (геликометра), 7 - базилярная мембрана, 8 - нижняя полость (барабанная лестница), 9 - круглое окно



Чтобы рассмотреть вторую причину нелинейности, необходимо еще раз вернуться к механизму преобразования сигнала на базиллярной мембране - механические смещения мембраны передаются органу Корти, это коллекция специальных нервных клеток, называемых волосковыми, расположенных рядами вдоль базиллярной мембраны, часть этих клеток называется внутренними (ВВК), их порядка 4000, другая часть - наружными (НВК), их около 12000 (рис. 3). Волосковые клетки - это механо-электрический преобразователь, который конвертирует механические смещения мембраны в электрический потенциал, что вызывает поток электрических импульсов (в двоичном коде) в связанных с ними нервных волокнах, т. е. они работают аналогично аналого-цифровому преобразователю.

В последние годы удалось установить, что ВВК связаны в основном с восходящими нервными волокнами, т. е. они, в основном, сообщают звуковую информацию в высшие отделы мозга - это "слуховые микрофоны", а НВК - с нисходящими нервными волокнами, т. е. они в основном получают приказы от мозга. Именно эти наружные волосковые клетки и играют основную роль в нелинейной компрессии звука. При больших уровнях сигнала они удлиняются (на 10% от основной длины) и, тем самым, как бы придерживают смещения базиллярной мембраны, предохраняя внутренние волосковые клетки от слишком большого изгиба, а на малых уровнях сигнала они усиливают смещения, как бы "подкачивая" энергию базиллярной мембране. Это было выявлено с помощью очень тонких современных экспериментов, позволивших обнаружить на очень низких уровнях сигнала отоакустическую эмиссию - т. е. излучение от внутреннего уха). Эта работа НВК на низких уровнях и вызывает, по-видимому, несколько anomальное поведение кубичных комбинационных тонов.

Измерения, выполненные с помощью анализа гамма-излучения от радиоактивного источника, размещенного на живой мембране, показали, что зависимость смещения базиллярной мембраны от уровня сигнала имеет вид нелинейной компрессии. Нужно отметить, что этот механизм работает только в определенном диапазоне сигналов, при очень длительном воздействии громких звуков, а это сейчас происходит очень часто, НВК повреждаются и эта зависимость приобретает чисто линейный характер, что приводит к дальнейшему разрушению ВВК и слухового нерва.

Таким образом, появление слышимых "слуховых" гармоник и комбинационных тонов является признаком нормальной работы слухового аппарата и свидетельствует скорее о хорошем состоянии слуха.

В заключение хотелось бы еще раз отметить, что в механизме слухового восприятия звука заложена нелинейная процедура обработки, обусловленная как гидродинамическими процессами в улитке, так и электромеханическими преобразованиями в волосковых клетках. Нелинейность слуха проявляется как при больших, так и при малых уровнях звукового сигнала и играет существенную роль в слуховом восприятии музыкальных, речевых и шумовых сигналов. Это полезно учитывать в практике работы музыкантов и звукорежиссеров.



Рис. 2а Вихревые токи в улитке



Рис. 2б Завихрения в воде вследствие удара о поверхность

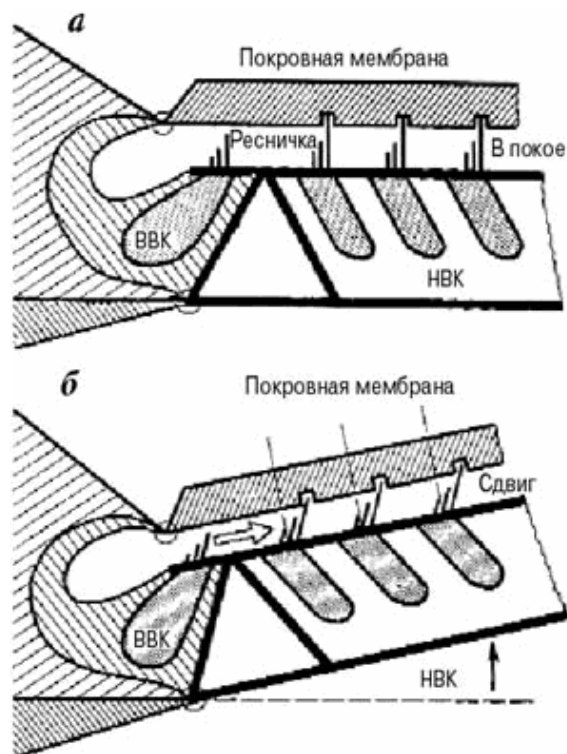


Рис. 3 Относительное положение в покое и в движении внутренних и наружных волосковых клеток

Часть 3 Слуховой анализ консонансов и диссонансов

Способность слуховой системы классифицировать звуки по высоте лежит в основе построения звуковысотных отношений в различных музыкальных культурах. Как уже было показано в первой части, для простых звуков определение высоты звука зависит в первую очередь от частоты сигнала, но также и от его интенсивности и длительности. Для сложных звуков это определяется способностью слухового аппарата делать спектральный анализ его состава, выделять и анализировать частотные соотношения между его гармониками и выявлять в нем признаки периодичности, так как только периодическим сигналам может быть присвоена высота, отнесенная к основному тону. Эта же способность слуховой системы к спектральному анализу и определению

частотных интервалов между гармониками лежит в основе ощущения "консонантности" или "диссонантности" звучания различных музыкальных интервалов и аккордов.

Консонанс (от французского слова *consonance*) - согласие (согласное звучание), соответственно диссонанс - несогласное, нестройное звучание. Эти понятия можно рассматривать с разных позиций: музыкально-психологических - "консонанс" ощущается как мягкое звучание, представляющееся выражением покоя, опоры, а "диссонанс" как раздражающее, беспокойное, являющееся носителем напряжения и движения. Чередование консонансов и диссонансов создает "гармоническое дыхание" музыки. В разных музыкальных культурах и в разные периоды времени отношение к консонансным и диссонансным интервалам было различным: если во время Пифагора к консонансным интервалам относили только октаву, квинту и кварту, а в 13 веке и терции перешли в разряд консонансных, то в музыке 20 века уже широко используются малые интервалы, которые раньше считались резко диссонансными (малая секунда). Вопросы использования и взаимодействия консонансных и диссонансных интервалов определяются учением о гармонии, которая также меняется в разные эпохи с изменением музыкальных вкусов.

К анализу консонансов и диссонансов можно подойти и с психоакустических позиций, то есть рассмотреть, как влияют на их восприятие частотные соотношения между гармоническими составляющими сложных музыкальных звуков. Эти психоакустические отношения являются общими и зависят только от внутренних свойств слухового аппарата. Сейчас, когда в руках звукорежиссеров и музыкантов имеются огромные возможности выбора различных интервалов и аккордов с помощью компьютерных технологий, кажется полезным рассказать об этом, чтобы при создании различных электронных композиций и обработке звукового материала в процессе звукозаписи учитывались особенности слуховой системы воспринимать определенные интервалы и аккорды как раздражающие (диссонансные) или наоборот.

Каждая нота, сыгранная на любом инструменте - это сложный звук, состоящий из основного тона и большого числа обертонов. Обертоном называется любая собственная частота выше первой, но только те обертоны, частоты которых относятся к частоте основного тона как целые числа, называются гармониками, причем основной тон считается первой гармоникой. Если этот звук дает четкое ощущение высоты тона, то он содержит в своем спектре только гармоники, то есть является периодическим (только периодические сигналы дают ощущение высоты тона).

Рис. 1. Отношения частот и музыкальные интервалы между первыми десятью гармониками натурального ряда тона Сз

Значения частоты каждой гармоники относятся к основному тону и друг другу как: $1f_0$, $2f_0$, $3f_0$, $4f_0$, $5f_0$, $6f_0$, $7f_0$

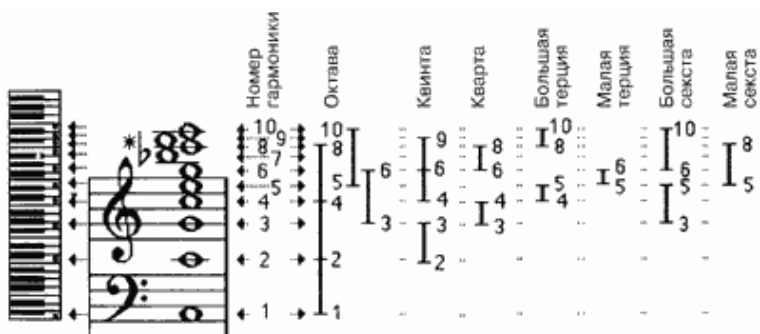
Если взять, например, за основной тон ноту до малой октавы и отложить от нее частоты с отношением 2:1, 3:1, 4:1, 5:1 и т. д., то мы получим обертоновый ряд, показанный на рисунке 1. Отношения частот гармоник друг к другу (они называются интервальными коэффициентами) также подчиняются отношению целых чисел и дают основные интервалы: 2:1-октава, 3:2-квинта, 4:3 -кварта, 5:4-мажорная терция и т. д. Музыкальные интервалы между гармониками уменьшаются по мере увеличения их номера в следующих пропорциях: $2:1 > 3:2 > 4:3 > 5:4 > 6:5$...

Каждая музыкальный тон теоретически содержит бесконечно большое число гармоник, соответствующих числу собственных частот колебаний струны, язычка и пр. Однако амплитуды их уменьшаются, и они практически становятся неслышимыми (всего попадает в слышимый диапазон, например, для ноты ля первой октавы $16000 \text{ Гц} / 440 \text{ Гц} = 36$ гармоник; если эту ноту сыграть на октаву выше, то в слышимом диапазоне остается 18 гармоник и т. д.)

Для многих инструментов имеются акустические пределы воспроизведения гармоник в силу механической природы их звучащего тела - у большинства акустических инструментов верхний предел лежит в пределах практического верхнего диапазона человеческого слуха до 16 кГц, хотя современные синтезаторы могут создавать сколь угодно большое число гармоник.

Как уже было показано в предыдущей статье, основное влияние на оценку высоты тона оказывают первые 7-8 "развернутых" гармоник, еще 8-9 гармоник несут дополнительную информацию как для оценки высоты, так и для оценки тембра звучания, то есть наиболее значимыми для слуха являются только первые 15-17 гармоник.

При оценке высоты тона производится спектральный анализ как с помощью оценки места максимального смещения на базилярной мембране, соответствующего данной частоте, так и с помощью оценки временных интервалов нейронных импульсов. Следует отметить, что распределение максимумов соответствует не самой частоте, а ее логарифму, именно поэтому слух одинаково оценивает интервал октава, если его образуют две частоты с отношением частот $200:100 \text{ Гц}$ или $2000:1000 \text{ Гц}$: по логарифмической шкале отношение этих двух расстояний одинаково и равно 2:1, по линейной - они отличаются в 10 раз. Поэтому практически при всех измерениях используется обычно логарифмическая шкала частот - это соответствует слуховому восприятию



интервалов.

Психоакустическая основа восприятия одних музыкальных интервалов как консонансных, других - как диссонансных, связана с понятием "критической полосы", которое имеет чрезвычайно большое значение как для восприятия мелодии и гармонии музыки, так и для современных систем сжатия звуковой информации в цифровом радиовещании и звукозаписи.

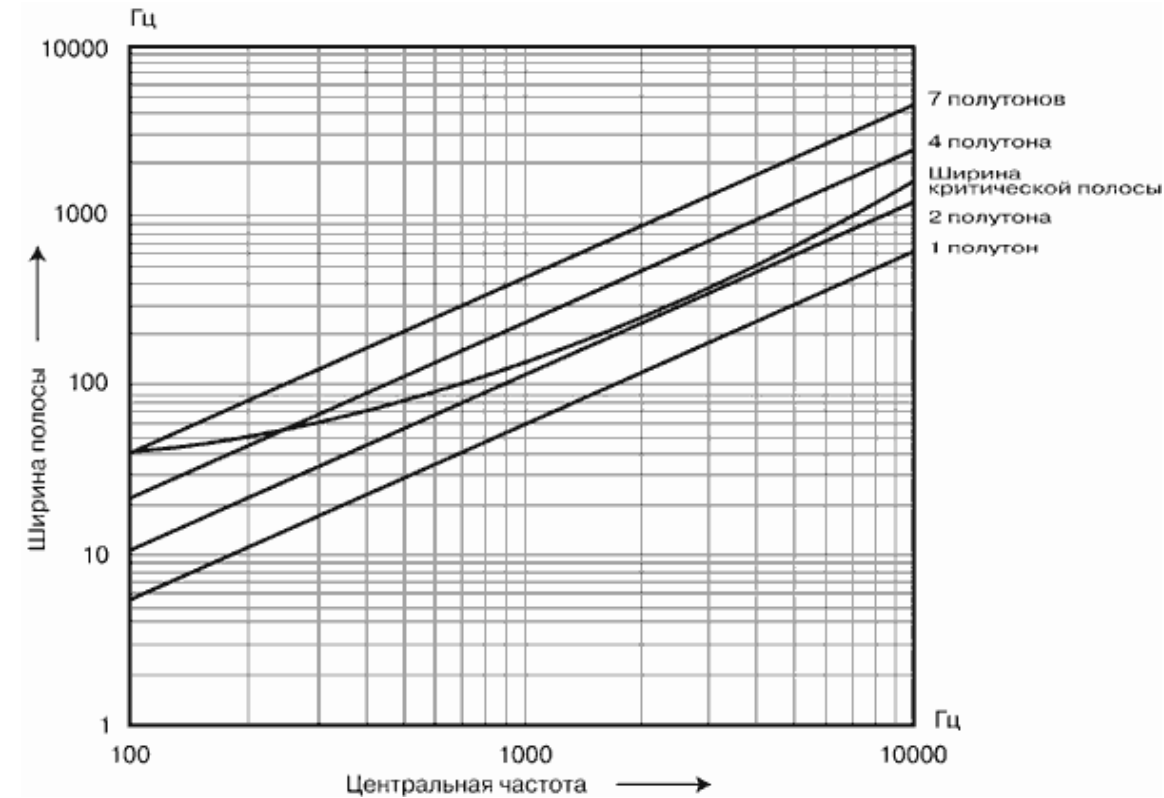


Рис. 2. Зависимость ширины критических полос от частоты

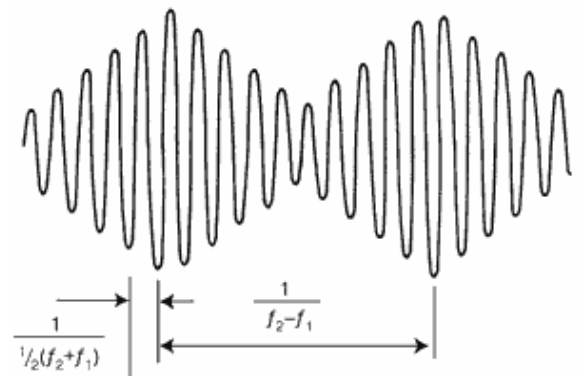
Действие базиллярной мембраны при спектральном анализе сложного звука можно считать эквивалентным действию линейки полосовых фильтров, каждый фильтр имеет ассиметричную форму с более крутым спадом в сторону высоких частот. Частотнозависимая ширина полосы пропускания фильтров зависит от разрешающей способности слуховой системы и определяет ширину "критической полосы". Определение "критической полосы" в современной литературе принято следующее: "ширина полосы, внутри которой слуховые ощущения резко изменяются". Действительно, ощущения громкости, маскировки и др. при попадании звуковых сигналов внутрь или вне критической полосы по частоте резко различаются. Зависимость ширины критических полос от частоты показана на рисунке 2 (для сравнения приведены линии, соответствующие ширине интервала в один полутона, два полутона, 4 и 7 полутонов на разных частотах). Из рисунка видно, что ширина критических полос с повышением частоты расширяется.

Следует понимать, что на базиллярной мембране действует подвижная линейка фильтров, при переходе от одних тонов к другим их центральные частоты меняются. Всего на базиллярной мембране размещается примерно 24 критических полосы с частотнозависимой шириной.

Ощущения диссонансности или консонансности созвучий также связано с наличием критических полос, то есть с конечной разрешающей способностью слуховой системы.

Рис. 3. Пример биений

Как известно из теории колебаний, если в системе происходит сложение двух колебаний с близкими частотами f_1 и f_2 , то возникает режим биений, эти биения воспринимаются на слух как пульсации громкости тона со средней частотой $1/2(f_1 + f_2)$ и медленно меняющейся амплитудой с частотой $(f_1 - f_2)$. Пример биений показан на рисунке 3. Когда частоты совпадают, два тона звучат в унисон, если начинать увеличивать частоту одного тона, то, вплоть до разницы 15 Гц, отчетливо прослушивается один тон с меняющейся громкостью - "биения", при дальнейшем увеличении разницы частот начинают прослушиваться оба тона с сильной шероховатостью звучания и, наконец, когда разница частот становится больше критической полосы - шероховатость исчезает.

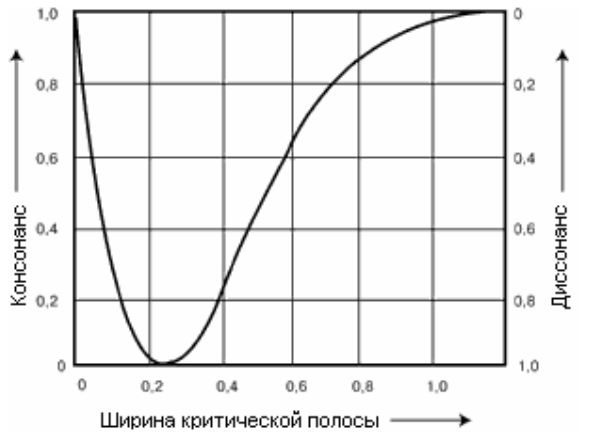


Это процесс можно легко прослушать, подав на акустическую систему два чистых тона от генератора, частота одного должна быть фиксирована, частота другого меняется.

Этим свойством, возникновением отчетливых биений, пользуются для настройки музыкальных инструментов. Частота F, на которой начинают прослушиваться два тона с сильной "шероховатостью", называется частотой "перемешивания". Она соответствует примерно разности частот около полутона, то есть $df/f = 0,06$ (на 500 Гц) и более чем целый тон $df/f = 0,12$ (на частотах ниже 200 и выше 4000 Гц).

Рис. 4. Зависимость степени ощущения консонантности (диссонантности) интервалов между двумя чистыми тонами в зависимости от ширины критической полосы

Эксперименты, сделанные с большой группой слушателей, среди которых не было профессиональных музыкантов (поскольку их слух натренирован на заученные образцы консонансных и диссонансных созвучий), позволили установить, при какой разнице по частоте два чистых синусоидальных звука воспринимаются как "приятные" консонансные или как резкие, неприятные "диссонансные". Результаты экспертиз были количественно обработаны и представлены на следующем графике (рисунок 4). Максимальная "приятность" звучания - консонанс - обозначен 1, диссонанс - 0, максимальная неприятность, "резкость" - консонанс - 0, диссонанс -1.



Как видно из графика, если разница частот равна нулю, то есть два тона звучат в унисон, то это совершенный консонанс. Если разница частот больше, чем критическая полоса, то это созвучие тоже звучит как консонанс. Для частот, разница между которыми составляет от 5 до 50% от критической полосы, созвучие воспринимается как диссонанс. Максимальный диссонанс прослушивается, когда разница составляет одну четверть от ширины критической полосы. Следует помнить, что ширина эта меняется с частотой (смотри рисунок 2). Поэтому два тона могут звучать как консонансный интервал в одной октаве, и как значительно менее консонансный (или даже диссонансный) - в другой.

Эти результаты полезно иметь в виду при составлении различных электронных музыкальных композиций и компьютерной обработке звука. Следует с осторожностью использовать сочетания звуков, частотная разница между которыми порядка одной четверти критической полосы - если не ставить специальной задачи создать такую музыку, чтобы слушатель от нее впадал в нервное расстройство.

Полученные результаты могут служить базой для определения степени консонансности различных интервалов и музыкальных аккордов сложных музыкальных тонов, содержащих в спектре большое количество гармоник.

В этом случае биения могут возникать как между фундаментальными частотами различных тонов, так и между их гармониками. Используя полученные выше результаты для простых тонов, можно количественно оценить степень консонансности (диссонансности) отдельных музыкальных интервалов.

В таблице 1 рассмотрены два тона, отношения фундаментальных частот которых равно 3:2, (квинта), нижняя частота 220 Гц.

Таблица 1

Первые семь гармоник нижнего тона, Гц	220	440	660	880	1100	1320	1540
Гармоника верхнего тона, Гц		330	660		990	1320	1650
Разница между частотами двух соседних гармоник, Гц	-	110	0		110	0	110
Средняя частота между гармониками, Гц		385	унисон		1045	унисон	1595
Ширина критической полосы, Гц		65	-		133	-	193,5
Половина ширины критической полосы, Гц		32,5			66,5		96,7
Степень консонантности/диссонантности (C, c, D, d)		c	C		d	C	d

Методика оценки степени консонанса (диссонанса) интервала в табл. 1 и табл. 2 основана на сравнении разницы частот двух соседних гармоник с шириной критической полосы, соответствующей средней частоте между ними:

Таблица 2

Первые семь гармоник нижнего тона, Гц	55	110	165	220	275	330	385
Гармоники верхнего тона, Гц	69,75		137,5	206,3	275	343,8	412,5
Разница между частотами, Гц	13,8		27,5	13,8	унисон	13,8	27,5
Средняя частота между гармониками, Гц	61,9		151	213	-	337	399
Ширина критической полосы, Гц	34,3		42,8	48,7	-	60,7	66,8
Половина ширины критической полосы, Гц	17,2		21,4	24,4	-	30,4	33,4
Степень консонантности/диссонантности (C, c, D, d)	D		d	D	C	D	D

- если две гармоники имеют равные частоты, или различие между ними меньше 5% от ширины критической полосы, то они обозначаются как совершенный консонанс - C;
- если разница между двумя гармониками по частоте больше ширины критической полосы (столбец 3 и 5), то это несовершенный консонанс - c;
- если разница между частотами ближайших гармоник меньше ширины критической полосы, то это диссонанс - d;
- если эта разница меньше половины ширины критической полосы, то это совершенный диссонанс - D.

Если частотная разница между большинством гармоник двух тонов больше ширины критической полосы или ее половины, то такое созвучие будет восприниматься как консонанс, поэтому, например, квинта относится к консонансным интервалам (рис. 5).

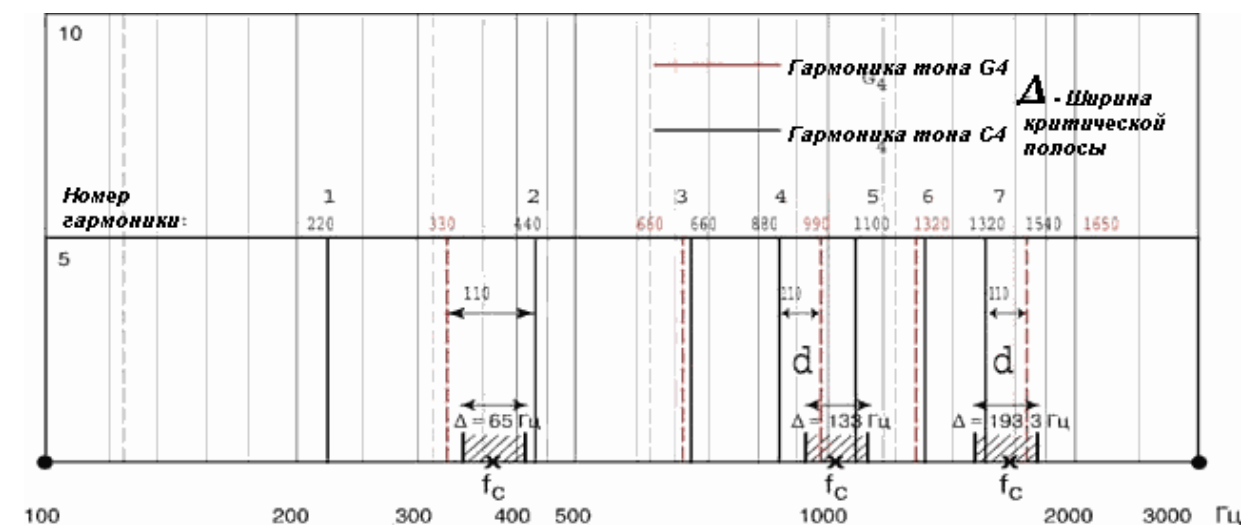


Рис. 5. Сравнение частотной разницы между соседними гармониками с шириной критической полосы

Приведем для примера результаты расчета для большой терции, отношение частот 5:4, нижний тон 55 Гц.

Как видно из полученных результатов, это интервал следует отнести к диссонансным.

Необходимо отметить, что один и тот же интервал или аккорд будет восприниматься как более или менее консонансный (диссонансный) в зависимости от того, в каком месте частотной шкалы он находится (так как ширина критической полосы частотно-зависима). Как следует из практики и подтверждается вышеприведенной методикой, уменьшающиеся интервалы между высокими гармониками (7:8, 8:9 и др.) звучат более диссонансно, чем интервалы между первыми гармониками (1:2, 2:3, 3:4 и др.). Решающую роль в слуховом ощущении степени консонантности (диссонантности) интервала играют развернутые первые 7-8 гармоник, как и при определении высоты тона.

Таким образом, способность слуховой системы воспринимать определенные сочетания звуков как благозвучные (консонансные) или раздражающие (диссонансные) связана с конечной разрешающей способностью слуховых фильтров и является ее фундаментальным свойством.

Ведущий российский специалист по акустике и аудиотехнологиям, доктор технических наук, профессор Ирина Аркадьевна Алдошина избрана членом Совета директоров AES (Audio Engineering Society), самого авторитетной и представительной организации в мире звуковой техники и технологии. Впервые представитель России вошел в руководство AES. Редакция журнала "Звукорежиссер" поздравляет Ирину Аркадьевну Алдошину, нашего постоянного автора, с избранием на высокий пост, и расценивает это событие как высокую оценку международной профессиональной общественностью ее трудов и заслуг.

Часть 4. Бинауральный слух и пространственная локализация

Ирина Алдошина

Наличие двух приемников слуха обеспечивает человеку возможность воспринимать пространственный звуковой мир и оценивать перемещение звуковых сигналов в пространстве. Информация, которая поступает на оба слуховых канала, обрабатывается в периферической части слуховой системы (подвергается спектрально-временному анализу) и затем передается в высшие отделы головного мозга, где путем сравнения этой информации из двух разных каналов формируется единый пространственный слуховой образ.

Восприятие через два приемника информации, иначе называемое бинауральным слухом, дает человеку огромные преимущества, основные из которых следующие:

- локализация сигналов как от одиночных, так и от множественных источников, что позволяет формировать

- пространственную перспективу и оценивать пространственное звуковое поле(например, в помещении).
- разделение сигналов,приходящих от различных звуковых источников из различных точек пространства.
 - выделение сигналов выбранного звукового источника на фоне других звуковых сигналов, например выделение прямого звука на фоне реверберирующих сигналов в помещении, выделение речи на фоне шумов и т.д.

Анализ бинауральных слуховых эффектов представляет особый научный интерес, в частности для изучения функционирования и спецификации полушарий головного мозга, а также громадный практический интерес в связи с развитием и промышленным внедрением бинауральных технологий для создания систем пространственной звукозаписи и звуковоспроизведения (стереофонические системы, пространственные системы типа Dolby Digital и др.), для синтеза трехмерных виртуальных звуковых полей (технология 3D-Sound,техника аурализации, создание адаптивных процессоров и др.), для развития новых методов метрологии и оценки звуковой аппаратуры.

Обеспечение

Рис 1б Огибание головы на низких частотах

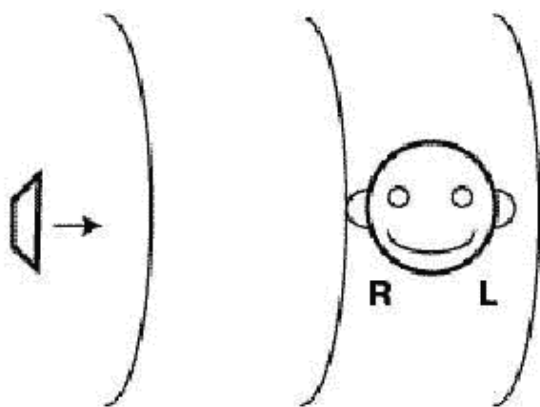
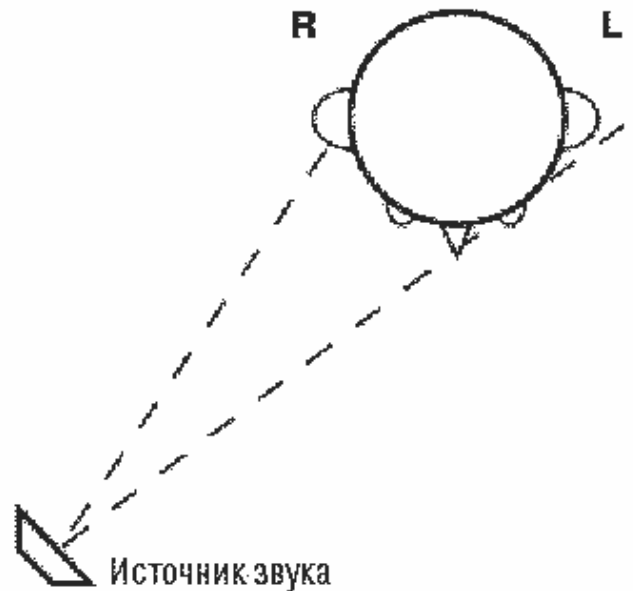


Рис 1а Расстояние от звукового источника до правого и левого ушей



пространственной панорамы, разделимости и выделения сигналов на фоне других сигналов и шумов является важнейшей задачей звукорежиссера при записи и обработке звука, а поскольку это требует использования бинауральных свойств слуха, то анализ этих свойств и является целью данной статьи.

К числу основных свойств бинаурального слуха можно отнести: пространственную локализацию, эффект предшествования, бинауральное суммирование громкости, бинауральную демаскировку, бинауральные биения и слияние звуков при определении высоты, эффекты "правого" и "левого" уха при восприятии речи и музыки и др.

Начнем рассмотрение этих свойств с пространственной локализации.

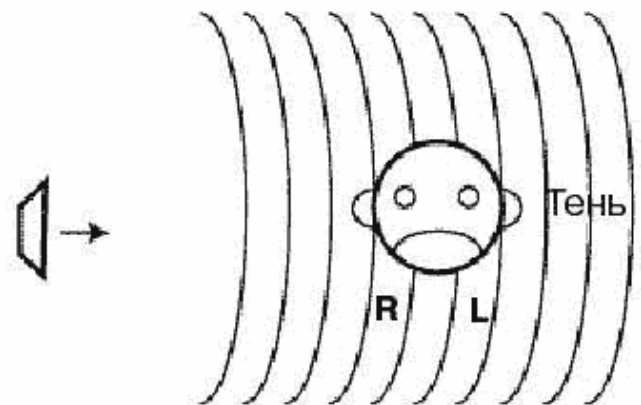
Бинауральная пространственная локализация

Прослушивая звучание симфонического оркестра в концертном зале (или пение хора в большом соборе), слушатель отчетливо воспринимает и разделяет расположение инструментов в горизонтальной плоскости на сцене, их расположение по глубине, а также ощущает пространственность окружающего звукового образа. Эта способность и называется пространственной бинауральной локализацией. Причем механизмы локализации в горизонтальной, вертикальной плоскости и по глубине несколько различаются.

Горизонтальная (азимутальная) локализация

На рисунке 1а представлены различительные признаки направленности при прослушивании источника звука (например, громкоговорителя при его различных положениях относительно головы слушателя). Звук, исходящий из громкоговорителя, расположенного справа от слушателя, должен пройти большее расстояние к левому уху, чем к правому. Как показано на рисунке 1б, низкие звуковые частоты имеют длину волны больше, чем диаметр головы, поэтому они огибают голову, поступая в ухо, расположенное дальше (дифракция). Однако звуки высокой частоты (Рис. 1в) имеют длину волны меньше, чем диаметр головы, поэтому они "блокируются"

Рис 1в Акустическая тень на высоких частотах



на пути к левому уху. Эта "акустическая" тень головы уменьшает интенсивность звука, поступающего в ухо, расположенное дальше от источника звука.

Частота, Гц	Длительность периода 712, град	(-)max, град	Частота, Гц	Длительность половины периода T/2, мс	(-)max, град
400	1,250	90	2000	0,250	24
800	0,625	90	2400	0,208	19
1200	0,417	42	3200	0,151	14
1600	0,313	30	4000	0,125	11

Пространственная разнесенность двух слуховых приемников (ушных раковин), и экранирующее влияние головы и торса за счет дифракционных эффектов приводит к значительным различиям между сигналами, поступающими в правое и левое ухо, что позволяет произвести локализацию звукового источника в пространстве, обусловленную тремя физическими факторами:

а) временным (Interaural Time Difference - ITD) - возникающим из-за несовпадения по времени моментов прихода одинаковых фаз звука к левому и правому уху;

б) интенсивностным (Interaural Intensity Difference - IID) - возникающим из-за неодинаковой величины интенсивностей звуковой волны вследствие дифракции ее вокруг головы и образования "акустической тени" со стороны, обратной источнику звука, как показано на рисунке 1а;

в) спектральным - возникающим из-за разницы в спектральном составе звуков, воспринимаемых левым и правым ухом, вследствие неодинакового экранирующего влияния головы и ушных раковин на низкочастотные и высокочастотные составляющие сложного звука.

а) временная разность - ITD

Разность времени прихода одинаковых фаз звука к ушам (ITD) можно легко рассчитать, зная разность хода dx звуковой волны до левого и правого уха $ITD=dx/C$, где C -скорость распространения звуковой волны.

Смысл величины ITD можно понять из рисунка 2. Исследования зависимости между направлением локализации источника звука в горизонтальной плоскости, определяемым углом $*$ и временем задержки ITD, приводят к следующему простому соотношению:

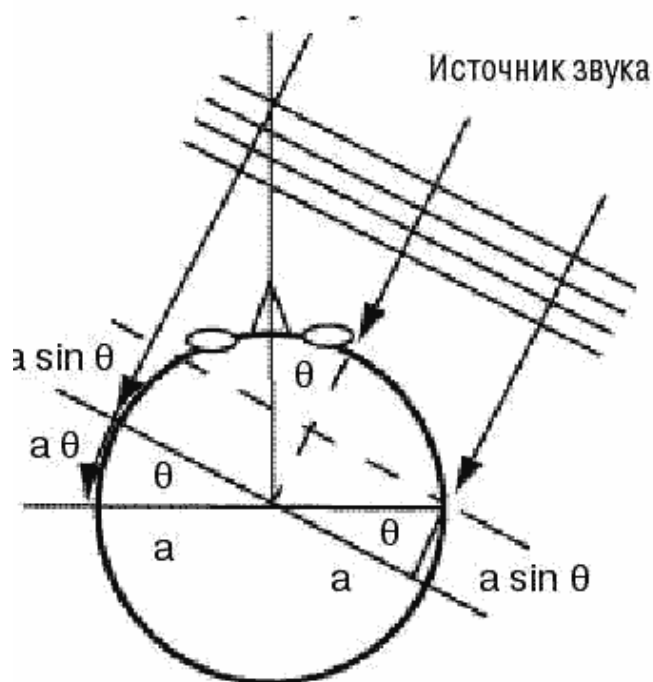
$$ITD=a/C (*+\sin *) \text{ при } -90^\circ < * < +90^\circ, (1)$$

где $*$ - азимутальный угол, отсчитываемый в горизонтальной плоскости от плоскости симметрии головы (Рис. 2); a - радиус головы. Разность времени прихода одинаковых фаз звука к ушам (ITD) равна 10^{-6} секунд при расположении звукового источника точно посередине и равна $a/c(*/2+1)$ для расположения источника точно напротив одного уха, что составляет $\sim 0,7$ мс (средний радиус головы ~ 9 см, кратчайшее расстояние вокруг головы от одного уха до другого ~ 26 см).

Различия по времени прихода звуковых волн для разных углов расположения источника для частоты 1500 Гц показаны на рисунке 3. Как видно из рисунка, при перемещении источника звука вокруг головы максимальная разница во времени возникает при $* = 90^\circ$. На низких частотах эта временная разница увеличивается.

Для синусоидальных колебаний при частоте 800 Гц максимальное время запаздывания ITD становится равным половине периода колебания $T/2$, а при более высоких частотах - превышает половину периода ($ITD > T/2$). В этом случае возникает неясность в фазовых соотношениях колебаний, действующих на правое и левое уши: с одинаковым основанием можно считать, что одна волна отстает по фазе от другой на время dT паеунд или опережает ее на это же время. Следовательно, предельное значение времени запаздывания, правильно воспринимаемое слухом, не должно превышать половину периода.

Рис 2 Разность хода лучей до левого и правого ушей



В соответствии с этим наибольшее значение азимутального угла *мах, определяемое временным бинауральным эффектом, с повышением частоты уменьшается. Это иллюстрируется данными табл. 1, в которой приведены расчетные значения *мах, вычисленные для разных частот по формуле (1) путем подстановки $ITD = T/2$. Например, при частоте 3200 Гц время запаздывания $ITD = T/2$ создает ощущение углового перемещения всего лишь на 14°. Однако это обстоятельство не столь существенно, так как в этой области частот при изменении направления прихода звуковых волн уже достаточно сильно сказывается дифракция звука вокруг головы, то есть вступает в силу интенсивностный фактор.

б) интенсивностная разность - IID

Как видно из рисунка 1а, по мере повышения частоты за счет дифракции образуется "акустическая тень" и интенсивность звуков, достигающих противоположного по отношению к источнику уха становится меньше. Наибольшая разность уровней звуковых давлений, действующих на левое и правое ухо, возникает при боковом положении источника (90°). Для этого случая на рисунке 4 приведен полученный экспериментально график частотной зависимости разности уровней звуковых давлений d N у левого и правого уха. Из графика видно, что по мере повышения частоты эта разность существенно возрастает, достигая на 5000 Гц величины ~20 дБ.

Последнее обстоятельство, однако, не означает, что при повышении частоты звука обостряется локализация. Напротив, чистые тоны очень высоких частот (свыше 8000 Гц) почти не поддаются локализации. Так же слабо выражена способность человека определять направление на источник синусоидальных звуков низкой частоты (ниже 300 Гц она становится значительно хуже, а ниже 150 Гц отсутствует вообще), поэтому в современных системах "домашний театр" расположение низкочастотных блоков (subwoofer) может выбираться произвольно.

Исследования ошибок при локализации положения синусоидального источника показали (Рис. 5), что наибольшие ошибки человек совершает в области 2000-4000 Гц, где, по-видимому, происходит смена механизмов локализации от временного к интенсивностному.

Анализ способности к угловому различию двух источников, находящихся в горизонтальной плоскости, также подтвердил, что в области частот 1500-2000 Гц резко возрастает наименьшая различимая величина угла между источниками.

Интересно отметить, что минимальное различие в азимуте (угле) воспринимается, когда источники находятся перед испытуемым. В этом случае он достигает 2i. Наибольшее различие возникает, когда источники находятся справа или слева: возникает так называемый "конус неопределенности" с каждой стороны уха (Рис. 6), внутри которого изменение положения источника звука не вызывает ощущение изменения его положения. Это объясняется тем, что при расположении источника сбоку получается большая разница и в интенсивности, и во времени, поэтому сдвиги источника дают малое относительное изменение общей разности. И поэтому для локализации очень важно движение головы - это изменяет положение конуса и сводит на нет его влияние.

с) спектральные различия

Рис 3 Различие междуушного времени при разных углах расположения источника звука

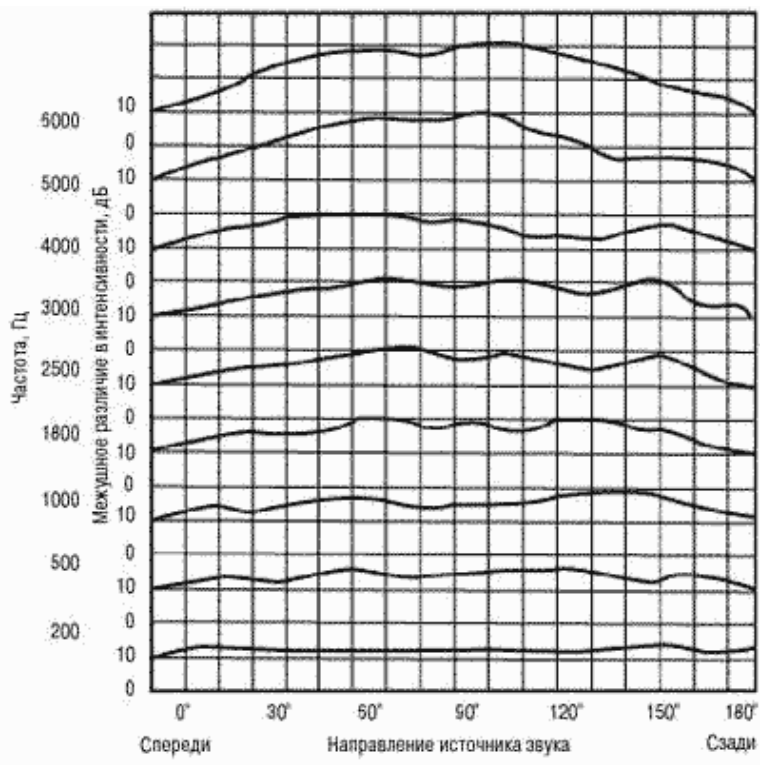
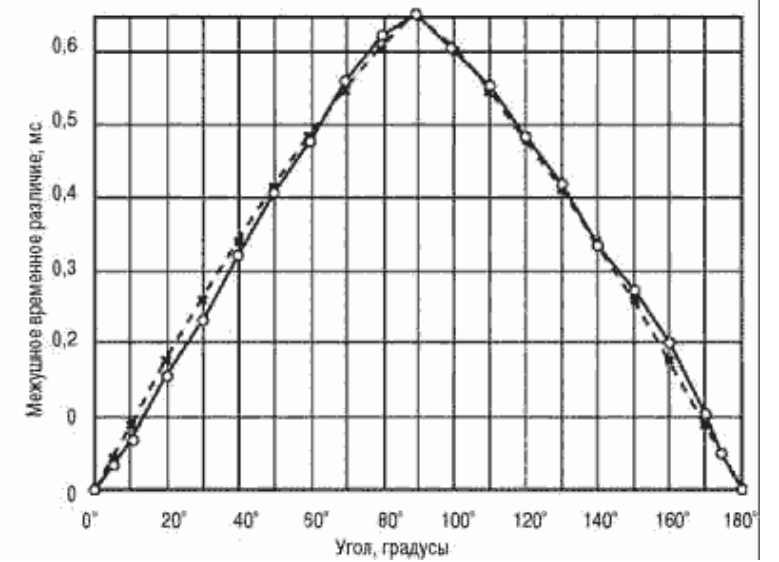
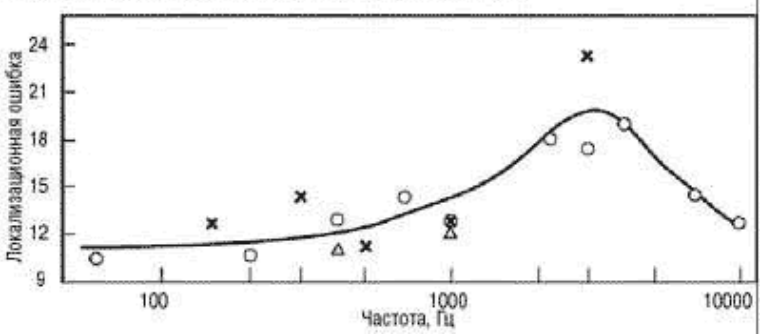


Рис 4 Межушные различия по интенсивности при разных углах

Рис 5 Локализационная ошибка как функция частоты



Наибольшая острота локализации достигается при восприятии сложных звуков и звуковых импульсов, когда, кроме рассмотренных ранее причин, сказывается еще и спектральный фактор. Например, если звук, приходящий под углом $\theta = 90^\circ$, содержит как низкочастотные, так и высокочастотные составляющие, то в спектре звука, действующего на дальнее ухо, высокочастотных составляющих будет меньше, так как на этих частотах скажется теневое действие головы.

Кроме того, сами ушные раковины производят сложную фильтрацию звука, зависящую от его частоты, что будет рассмотрено дальше. Существенное значение для локализации имеет также энергия переходных процессов, причем наибольшее значение имеет наличие в звуке низкочастотных составляющих переходного процесса. Поэтому при прослушивании музыкальных и речевых сигналов изменение спектрального состава сигнала, а, следовательно, и его тембра, в зависимости от его расположения, помогает в локализации.

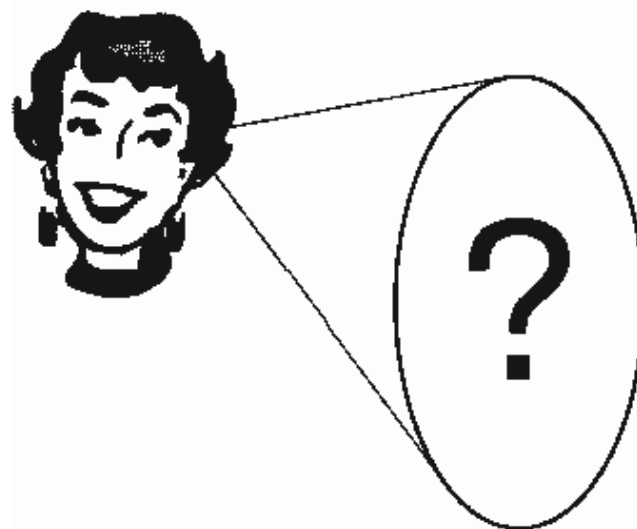


Рис 6 Конус неопределенности

В целом анализ способности к локализации в горизонтальной плоскости показал, что наименьший ощутимый угол отклонения источника при восприятии звуковых импульсов составляет около 3° . Эту величину следует считать угловой, или бинауральной разрешающей способностью слуха. Однако слух замечает угловое смещение на 3° , но при определении направления совершает ошибку в среднем на 12° . Поэтому точность локализации имеет величину 12° для источников, находящихся в передней полуплоскости, а для источников, расположенных позади слушателя, эта точность еще меньше.

Вертикальная (высотная) локализация

Способность определять направление прихода звука в вертикальной плоскости у человека развита значительно слабее, чем в горизонтальной. Она составляет $10-15^\circ$ (по сравнению с 3° в горизонтальной). Эту способность связывают обычно с ориентацией и формой ушных раковин: если в ушной канал поставить микрофоны и записать звук от источника, находящегося в разных точках медианной плоскости (также и в горизонтальной плоскости), то АЧХ (Рис. 7) будет разной при приходе звука спереди - сверху и сзади на АЧХ отчетливо видны пики за счет отражения от ушной раковины в области 4 - 8 кГц, хотя есть пики и ниже 2 кГц за счет отражения от грудной клетки и спины слушателя.

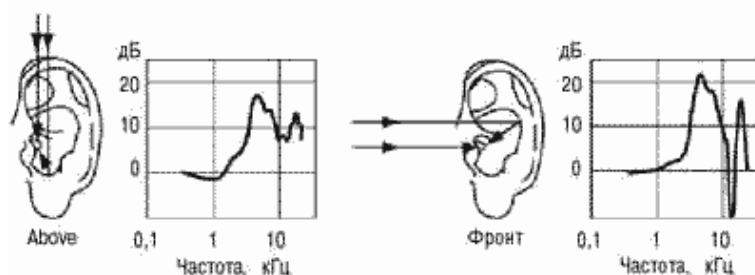
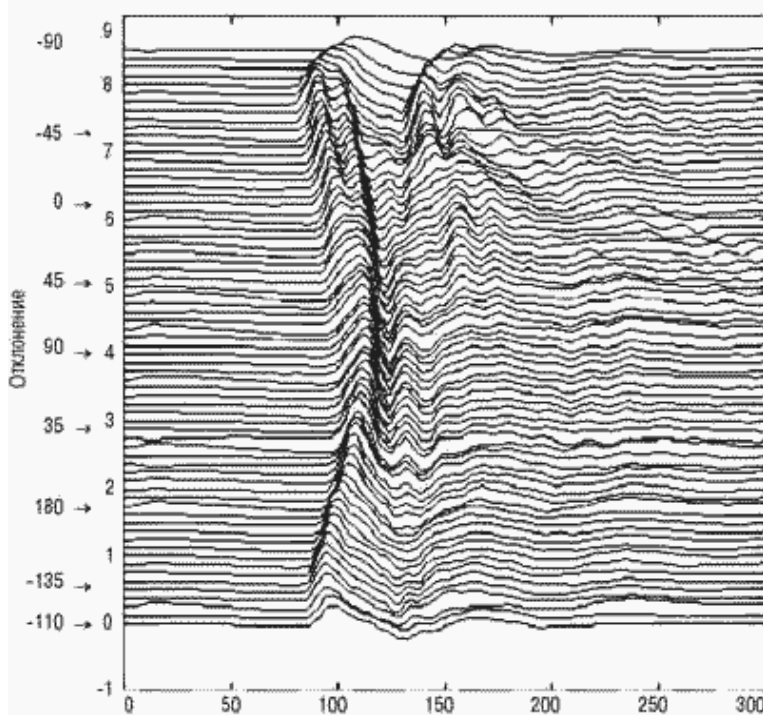


Рис 7 Форма АЧХ при разных углах расположения звукового источника

Ушная раковина имеет сложную геометрию, она действует как акустическая антенна: на низких частотах она усиливает общую энергию сигнала, на средних и высоких частотах начинают сказываться резонансы ее внутренних полостей, усиливая некоторые частоты. Кроме того, происходит интерференция прямого звука со звуком, отраженным от отдельных участков ушной раковины, то есть ушная раковина действует как фильтр, внося максимальные искажения в области 6-16 кГц, причем форма этих искажений зависит от того, спереди или сзади находится источник звука, и под каким углом подъема он расположен в медианной плоскости. Вид АЧХ сигнала, записанного на микрофоны, находящиеся в ушных раковинах при разных положениях источника, показан на рисунке 8 (они называются бинауральными передаточными функциями головы - HRTF).

Эта зависимость АЧХ звукового давления, поступающего на барабанную перепонку левого и правого уха, от положения источника, используется для сравнения спектральных компонент сигнала, приходящего спереди и сзади или и сверху, и их локализации. Поэтому широкополосные сигналы

Рис 8 Передаточные функции головы и ушных раковин



лучше локализируются, чем узкополосный шум.

Если звуковые сигналы подавать через наушники, то ушные раковины оказываются прижатыми к голове. Поскольку такая ситуация для мозгового процессора является неестественной, человек теряет способность производить локализацию в пространстве, помещая при этом источник звука как бы внутрь головы. Это свойство называется латерализацией и служит причиной значительной утомляемости людей, долгое время работающих в наушниках. В настоящее время созданы цифровые процессоры, которые производят предварительную фильтрацию сигналов в наушниках, аналогичную тому, как это делает ушная раковина. Это дает возможность "выносить" звуковой образ из головы, облегчая работу звукорежиссеров, операторов и др.

Глубинная локализация (оценка расстояния до источника)

Чувствительность слуха к расстоянию до источника имеет жизненно важное значение - гудок автомобиля, находящегося сзади близко или далеко, должен вызывать разную реакцию. Однако именно это свойство слуховой системы изучено явно недостаточно. Среди основных факторов, определяющих оценку глубины можно выделить следующие:

- уменьшение уровня звукового давления с расстоянием - на низких частотах, где длина волны большая (* *5-15 м), любой источник можно считать точечным, и звуковые волны вокруг него - сферическими. В сферической волне площадь поверхности увеличивается пропорционально квадрату расстояния, и соответственно давление падает обратно пропорционально расстоянию, то есть на 6 дБ при каждом удвоении расстояния.

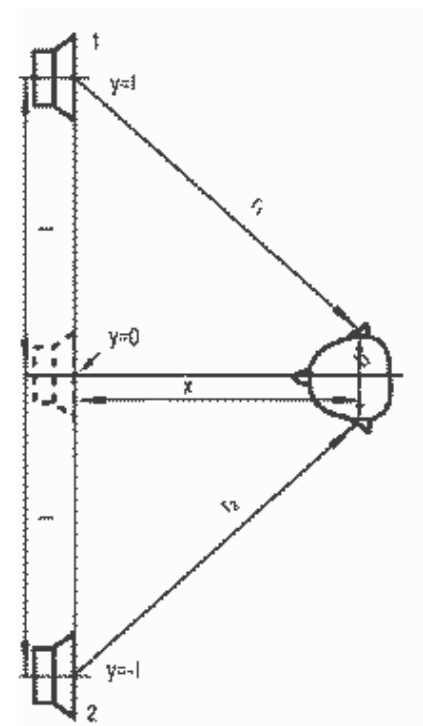


Рис 9 Расположение мнимого источника

Многочисленные эксперименты по смещению источника и оценке кажущего расстояния до слухового образа (выполненные в заглушенной камере и на открытом пространстве) показали, что, при удалении источника-громкоговорителя на расстояние от 1 до 10 м, слуховой образ у экспертов (в заглушенной камере при отсутствии визуального контроля) также смещался в этом же направлении, но имело место отставание слухового образа от реального источника - чем дальше, тем больше.

Ощущение удвоения расстояния до звукового объекта возникало только при уменьшении уровня звукового давления на 20 дБ (а не на 6 дБ, как при объективном измерении). При этом точность локализации была не очень велика: ошибка для широкополосного сигнала (щелчки, часы и т.п.) составляла от 3,5 до 30 см при изменении расстояния от 1 до 8 м. Если при увеличении расстояния повышать напряжение на громкоговорителе так, чтобы уровень звукового давления у слухового канала эксперта не менялся, то способность определять расстояние до источника (глубинная локализация) исчезает.

Таким образом, при отсутствии визуального контроля в условиях свободного поля, когда отраженные сигналы поглощаются (например, в заглушенной камере или в свободном пространстве), уровень звукового давления в месте расположения эксперта является решающим признаком, по которому и оценивается расстояние до источника.

При больших расстояниях (больше 15 м) начинает сказываться затухание, зависящее от расстояния, проходимо звуковой волной. При этом высокочастотные составляющие затухают быстрее, и спектральный состав сигнала при удалении источника меняется (тембр становится "темнее"). Кроме того, на распространение звука оказывает влияние влажность воздуха и направление ветра на открытом пространстве.

Следует отметить, что возможности слуха по определению глубины расположения источника ограничены, имеется "акустический горизонт".

На близком расстоянии (менее 3 м), на глубинную локализацию начинает оказывать влияние также дифракция на ушной раковине и голове, то есть сказываются разности уровней интенсивностей (выше 1500 Гц) и временные задержки (ниже 1500 Гц), как и в предыдущих случаях.

Приблизительно локализацию по глубине при расстояниях меньше 3 м можно оценить по формуле:

$L=2C \cdot dT (\ln I_0 / I)$, где dT - временная разность сигналов, I - интенсивность.

При этом на близких расстояниях меняется спектральный состав при смещении звукового источника за счет дифракционных эффектов, то есть меняется тембр ("тускнеет" при приближении к источнику).

Таким образом, при изменении расстояния до источника меняется одновременно громкость и тембр, что и служит различительными признаками.

Общая точность глубинной локализации не очень велика, при смещении широкополосного звукового источника от 50 до 150 см ошибки составляют 15-30%.

Существенную роль для глубинной локализации играет личный опыт, если слушателю знаком сигнал, а если

он имеет возможность сделать визуальную оценку, то точность глубинной локализации многократно увеличивается.

Точность глубинной локализации звукового источника значительно повышается в закрытом реверберирующем помещении. Роль реверберации в оценке удаленности источника, например, распределения музыкантов по глубине оркестра, исключительно велика. При перемещении звукового источника по глубине меняется отношение энергии прямого звука к энергии отраженного (реверберационного) звука, что помогает точнее определить расстояние до источника. Важнейшее значение имеет также разность по времени между прямым звуком и приходом первых отражений и соотношение их по уровням.

Приближенно, глубинную локализацию в помещении можно оценить следующим образом:

где α - коэффициент поглощения, S - площадь поверхности, $E_{\text{пр}}/E_{\text{пр}}$ - отношение плотностей отраженной и прямой энергии.

$$L = \frac{\alpha S}{50(1-\alpha)} \sqrt{\frac{E_{\text{пр}}}{E_{\text{пр}}}}$$

Субъективное ощущение "акустики зала" определяется целым рядом параметров, некоторые из них прямо связаны с пространственной локализацией:

Пространственное впечатление (камерность, интимность, близость) - определяет для слушателя кажущийся размер пространства. Разные стили музыки требуют разных его значений. Композитор (звукорежиссер, исполнитель и др.) должен иметь в виду этот параметр, иначе будет несоответствие стиля музыки размеру помещения (например, звучание органа в маленькой комнате), что очень четко ощущается слушателями.

Пространственное впечатление определяется разницей во времени между прямым звуком и первыми отражениями. В залах с "интимной" акустикой эта разница составляет для слушателей в центре зала 15-30 мс. Если эти отражения имеют похожие спектр и огибающую, и их громкость не выше прямого звука, то в пределах этого времени они не воспринимаются как отдельные отражения, а помогают в улучшении локализации прямого звука, в том числе глубинной. Малая разница во времени прихода первых отражений характерна для музыкальных комнат XXVIII столетия, средняя - для концертных залов 19 века, большая - для соборов.

Амбиентность - ощущение слушателя, что музыка от источника (например, оркестра) идет от всего фронта сцены, и звук окружает его со всех сторон.

Тренированный слушатель различает две составляющие в восприятии амбиентности: кажущееся расширение площади источника звука, и окружение (обволакивание), когда слушатель чувствует себя погруженным в звук, окруженным им со всех сторон.

По мнению многих экспертов, кажущееся расширение площади источника является одним из главных индикаторов акустического качества концертных залов и помещений прослушивания. Оно связано с уровнем боковых отражений - чем выше этот уровень, тем больше кажущееся расширение источника.

Кроме того, высокую связь с этим параметром показали результаты измерения на искусственной голове коэффициента внутрислуховой кросс-корреляции сигнала, усредненного в интервале 0-80 мс и измеренного в третьооктавных полосах с центральными частотами 500, 1000, 2000 Гц. Значения этого коэффициента (в соответствии с измерениями Беранека) для девятнадцати лучших залов мира составляют от 0,35 до 0,6. Кажущаяся ширина звукового источника связана также с уровнем звукового давления на низких частотах, в основном в области частот 125 и 200 Гц.

Обволакивание (окружение) - связано с ощущением позднего реверберирующего звука, поступающего со всех сторон (после 80 мс). Оно определяется конструкцией зала: наличием нерегулярностей стен, балконов и пр., то есть всеми конструктивными элементами, которые обеспечивают приход звука с разных сторон. Ощущения от звучания музыки у слушателя, к которому отраженные звуки приходят со всех сторон: от потолка, стен, пола и т.д., будут существенно отличаться от ощущений слушателя, сидящего под балконом, к которому звук приходит только с фронта. Оно связано с коэффициентом внутрислуховой кросс-корреляции, усредненного за период времени от 80 мс до 1 с.

Таким образом, наш слуховой аппарат, используя разные механизмы обработки звуковых сигналов, позволяет определить и локализовать положение звукового источника в трехмерном пространстве. Именно эта способность используется при создании современных систем компьютерного моделирования трехмерных звуковых пространств (системы аурализации).

Это же свойство слуха используется и в современных системах пространственного звуковоспроизведения. Создавая искусственные условия, к которым наша слуховая система не была приспособлена в процессе естественной эволюции, например, помещая два одинаковых громкоговорителя на одинаковом расстоянии от левого и правого ушей, подавая на них одинаковые сигналы, (Рис. 9), мы заставляем наш слуховой аппарат помещать слышимый (мнимый) источник звука посередине между реальными звуковыми источниками. Пространство таких мнимых источников, создаваемых различными пространственными системами воспроизведения (стереофоническими, Surround и др.), и создает стереоэффект - по существу, это "большой обман" нашего слухового аппарата. Вопрос о том, как формируется и как управляется этот пространственный образ мнимых (виртуальных) источников, может служить предметом рассмотрения отдельной статьи.

Часть 5. Бинауральный слух (продолжение)

Как уже было отмечено в предыдущей статье, кроме эффектов пространственной локализации, наличие бинаурального слуха, то есть двух слуховых приемников, обеспечивает целый ряд других преимуществ в получении и переработке слуховой информации.

К их числу можно отнести: бинауральную чувствительность и суммацию громкости, бинауральные слияния звукового образа и биения, эффект предшествования, бинауральную маскировку и демаскировку, эффекты "правого" и "левого" уха при восприятии речи и музыки и др.

Каждое из этих свойств слуха имеет огромное значение для восприятия окружающего нас звукового пространства и все в большей степени используется в современных звуковых технологиях записи, передачи и воспроизведения, особенно с помощью быстро развивающихся компьютерных методов обработки звука.

Остановимся в данной статье на первых трех свойствах бинаурального слуха, поскольку анализ двух последних требует дополнительных сведений о законах маскировки (на которых постараемся остановиться в дальнейшем).

Суммация звуков при бинауральном слухе

Анализ порогов слышимости, выполненный при моноуральном слушании и при бинауральном показал, что уровень слуховых порогов при бинауральном восприятии сигналов (синус, речь, шум, музыка) ниже, чем при моноуральном. Интенсивность звука для достижения порога слышимости при восприятии звука двумя слуховыми приемниками ниже на 3 дБ, то есть нужно создать в два раза больше акустическую мощность, чтобы звуковой сигнал, находящийся на пороге слышимости при прослушивании бинаурально, услышать при переходе на моноуральное прослушивание (одним ухом).

Таким образом, наличие двух слуховых приемников позволяет услышать значительно более тихие звуки, что имеет существенное значение для оценки окружающего звукового пространства.

Бинауральная суммация громкости проявляется в том, что, как показали эксперименты Флетчера, сигнал при заданном уровне громкости, например, 70 дБ, будет звучать в два раза громче, если он подается на два уха, чем на одно, то есть громкость удваивается (суммируется).

Построенные на разных частотах кривые зависимости оцененной громкости (сон) от уровня подаваемого сигнала показали, что по мере повышения уровня подаваемого сигнала преимущества бинаурального слуха возрастают: при уровне сигнала ниже 35 дБ, чтобы звуки были равногромкими при моно- и бинауральном слушании, подаваемый на два уха сигнал может быть на 3 дБ ниже по интенсивности. При уровне выше 35 дБ эта разница увеличивается, и остается примерно постоянной при дальнейшем увеличении уровня подводимого сигнала.

Дифференциальная чувствительность (то есть способность замечать различия в звуках, как по частоте, так и по интенсивности), как показали многочисленные эксперименты, при бинауральном слушании выше, чем при моноуральном.

Результаты, полученные в различных исследованиях, позволяют считать, что при бинауральном слушании дифференциальная чувствительность по интенсивности выше в 1,65 раза, по частоте выше в 1,44 раза.

Таким образом, наличие двух слуховых приемников позволяет услышать более тонкое различие звуков по высоте и по громкости, что имеет принципиально важное значение как для аудиотехники, так и для восприятия музыки.

Бинауральное слияние звуков и биения

Несмотря на то, что в обычных условиях в оба уха звуки поступают с определенным различием во времени, по интенсивности и спектру, мы воспринимаем один слуховой образ. Мы воспринимаем один мир двумя ушами. Точнее, в оба уха поступают подобные, но не идентичные звуки, сливающиеся в единый образ. Этот процесс носит название бинаурального слияния.

Слуховая система воспроизводит бинауральное слияние в течение всего времени подачи в оба уха звуков, сходных в определенном отношении (однако совершенно разные звуки не сливаются).

Наиболее важным для бинаурального слияния являются звуки с частотой ниже 1500 Гц. Эксперименты показали, что если подавать через наушники два высокочастотных звука с разными частотами, то они воспринимаются как отдельные звуковые сигналы, однако если эти сигналы промодулировать каким-либо низкочастотным звуком, то оба сигнала сливаются в единый слуховой образ.

Полученный результат свидетельствует о том, что для бинаурального слияния слуховая система использует низкочастотную огибающую комплексного звука (его макроструктуру), несмотря на то, что детали составляющих комплексного звука (его микроструктура) различны.

Бинауральное слияние речи, например, выявляется, когда в одно ухо поступают только высокочастотные компоненты речевого звука, а в другое - только низкочастотные. Несмотря на то, что ни одно ухо не получает

достаточной информации для распознавания речевого сигнала, получаемый в результате бинаурального слияния слуховой образ позволяет понять речь.

Бинауральное слияние может быть показано на эффекте "пропущенной фундаментальной" (о котором мы говорили в первой статье). При бинауральном прослушивании он может иметь место даже в том случае, если мы на одно ухо подаем четные гармоники: 200 Гц, 400 Гц, 600 Гц..., а на другое - нечетные: 300, 500, 700..., все равно будет идентифицироваться одна высота основного тона (в данном случае соответствующая 100 Гц).

Механизм бинаурального слияния звуков описан в виде математической модели, которая основывается на поиске центральной слуховой нервной системой перекрестных корреляций между звуковыми сигналами в обоих ушах. Другими словами, звуки, поступающие в уши, рассматриваются как статистические события, а механизм бинаурального слияния использует поиск общности между ними. Этот же процесс позволяет выделять периодические компоненты сигналов из шума, что важно для расширения динамического диапазона воспринимаемых звуковых сигналов при бинауральном слушании.

Когда один тон подается в правое ухо, а другой, незначительно отличающийся по частоте, - в левое, в слившемся слуховом образе воспринимаются биения, которые лежат в основе определения консонансных и диссонансных интервалов звуков. Интересная особенность бинауральных биений состоит в том, что они проявляются при полной акустической изоляции обоих звуков, поступающих в левое и правое уши. Очевидно, бинауральные биения возникают в определенном месте центральной нервной системы при взаимодействии нейронной активности, кодирующей поступающие в оба уха звуки. Нейроны, дающие ответную реакцию на огибающую бинауральных биений, обнаружены в нижних отделах головного мозга (на рисунке Superior olive (В)).

Бинауральные биения отличаются от моноуральных некоторыми особенностями: в то время как моноуральные биения могут быть слышимы при взаимодействии тонов всего воспринимаемого диапазона частот, бинауральные биения связаны с низкими частотами, и наибольшие бинауральные биения воспроизводятся при взаимодействии звуков с частотой от 300 до 600 Гц. Кроме того, бинауральные биения воспринимаются при существенной разнице в интенсивности между звуками, подаваемыми в оба уха, даже в случае, когда один из звуков подается на подпороговом уровне его интенсивности. Как уже было показано в предыдущей статье, биения возникают, когда разность частот обоих подаваемых звуков находится в пределах до 15 Гц.

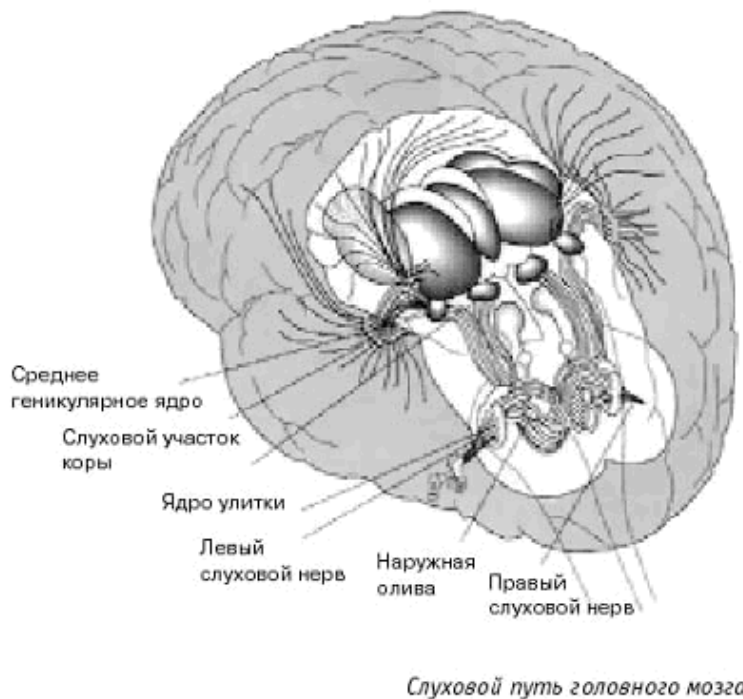
Неожиданное применение нашла способность слуха различать бинауральные биения в создании так называемых "генераторов мозговых волн" (brain wave generator). Если подобрать разность частот двух сигналов поступающих в оба уха через наушники, совпадающих с альфа-, бета- и другими ритмами мозга, то можно, по мнению авторов, улучшить сон, память и др. (подробнее об этом можно узнать в Интернете <http://www.bwgen.com>).

Эффект предшествования (эффект Хааса)

Анализ этой проблемы - одна из старейших тем в исследованиях бинаурального слуха. Эффект предшествования впервые детально описан в 1949 г., хотя о нем было известно и раньше.

В общем виде эффект предшествования заключается в том, что в пределах определенного отрезка времени ранее поступивший звуковой сигнал (фронт звуковой волны) доминирует в слуховом восприятии над звуками, поступившими позднее (эхо).

Рассмотрим, например, ситуацию, когда две акустические системы воспроизводят одинаковый сигнал одного уровня. Если слушатель находится на определенном расстоянии от них на средней линии, то в этом случае звук исходит из мнимого источника, находящегося между ними. Однако, если ввести задержку во вторую акустическую систему, то звук начнет перемещаться в сторону первой акустической системы. Как показал Хаас, при изменении задержки от 0 до 10 мс мнимый источник переместится и совпадет с первой акустической системой. При изменении задержки на второй акустической системе от 10 до 30 мс, звук будет казаться исходящим только из первой акустической системы (хотя вторая система будет продолжать воспроизводить звук той же интенсивности), то есть локализация будет производиться только по опережающему сигналу - в этом и состоит эффект Хааса. Звук второй системы как бы подавляется мозгом, хотя собственно слуховая



Слуховой путь головного мозга

система продолжает его слышать. Однако звук, приходящий от второй акустической системы, создает определенные ощущения объема.

При дальнейшем увеличении задержки от 30 до 50 мс, слушатель ощущает, что звук идет и из второй системы, хотя локализация продолжает идти на первую. Только при задержке более ~50 мс (это зависит от характера сигнала - речь, музыка и др.), ощущается звук второй системы, как эхо.

Разумеется, эти эффекты зависят от соотношения интенсивностей сигналов, от степени их подобия и их спектрального состава.

Это свойство бинауральной слуховой системы имеет огромное значение для оценки акустики помещения. В любом помещении слушатель воспринимает прямой звук от источника сигнала (певца, музыканта, лектора и др.) и отраженные звуки от стен помещения. Отраженные звуки поступают в уши позже, и будут иметь другое направление, чем прямой звук. Источник звука в этом случае локализуется по направлению прямого звука, а не отраженного. Хотя отраженные звуки и будут окрашивать, качественно изменять слышимый звук, восприниматься будет только ранее прибывший прямой звук. Сказанное применимо к отраженным звукам, поступившим только в определенном отрезке времени после поступления прямого звука.

В реверберационном процессе можно выделить два отрезка - "ранние" дискретные отражения до 80 мс (в зависимости от типа помещения), и "поздние" отражения со временем запаздывания больше 80 мс. Эффект предшествования подавляет ранние отраженные звуки, они интегрируются с прямым звуком в единый слуховой образ, сохраняя локализацию на источник прямого звука. Однако отраженные звуки вносят свою окраску в воспринимаемый звук, они несут информацию о пространственности, интимности, ясности и других субъективных параметрах, играющих решающую роль в оценке качества звучания в помещениях. Это показали работы известного акустика Беранека, выполненные им на протяжении многих лет в лучших залах мира. Отраженные звуки имеют важное значение для определения разборчивости речи в помещениях.

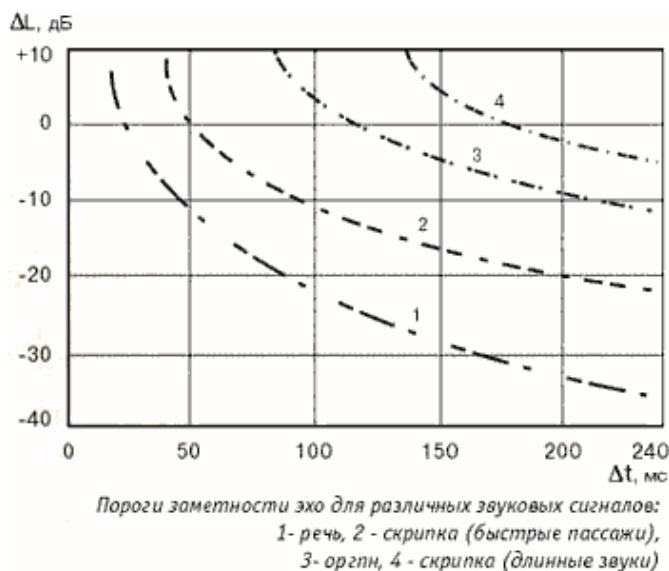
Отраженные сигналы в помещении могут восприниматься и как отдельные повторяющиеся сигналы - эхо, при этом уровень их осознанного восприятия зависит от времени задержки, соотношения их интенсивностей с прямым звуком, спектрального состава сигнала, степени заполнения паузы между приходом отраженных сигналов и др. Наличие эхо-сигналов в помещении оказывает отрицательное влияние на качество звучания музыки и разборчивость речи. Взаимосвязь порогов заметности эха от времени запаздывания и интенсивности отраженных звуков для разных сигналов (речи, скрипки, органа) показана на графике.

Наиболее низкими пороги оказываются для речи: чтобы отраженные сигналы не ухудшали разборчивость речи, необходимо, чтобы при задержке 50 мс они были ниже по уровню основного сигнала на -10 дБ, при 100 мс на -20 дБ и т.д., поэтому для повышения разборчивости речи необходимо обеспечивать высокий уровень прямого звука. Существенное влияние на пороги заметности эха оказывает спектр запаздывающих сигналов: исследования показали, что порог эха при высокочастотных сигналах ниже, чем при низкочастотных. При высокочастотных шумах, а еще в большей степени при высокочастотных импульсах, направление прихода звука распознается по бинауральной разности времени. В таких случаях (начиная с частоты 1,6 кГц) сравниваются, по-видимому, изменения огибающих сигнала за малые интервалы времени.

Наконец, влияние на пороги заметности эха оказывает направление прихода отраженных звуков: оценка мешающего влияния отраженных сигналов на речевой сигнал показала, что при боковом падении звука порог эха на 5 дБ ниже чем при фронтальном. Все эти данные особенно важно учитывать при построении систем звукоусиления, т.к. иначе это может привести к появлению сильных эхо-сигналов и потере разборчивости речи.

В помещениях, не имеющих сильных концентраций отражений, правильная локализация на источник звука благодаря действию эффекта предшествования сохраняется, даже когда энергия отражений превышает энергию прямого звука (до определенных пределов ~10дБ). Появление мешающего эха следует рассматривать как границу возможностей использования эффекта предшествования (первой волны).

В 1987 г. были опубликованы исследования Клифтона, который показал, что этот эффект является динамическим, и требует определенного времени для "обучения" слуховой системы: если в заглушенной камере установить два громкоговорителя и подать на них два коротких импульса, следующих друг за другом, то в первый момент времени слушатель воспринимает их как отдельные щелчки, затем (при повторении их со скважностью 10-12 периодов в секунду), восприятие второго импульса ослабевает и становится слышен только один импульс от первого громкоговорителя, а второй добавляет только некоторую объемность. Интересно, что если сделать небольшую паузу и повторить эксперимент, то слушатель сразу слышит один



звуковой образ от первого громкоговорителя. Можно предположить, что слуховая система за период "обучения" строит определенную модель акустического пространства, создавая таким образом основу для распознавания прямых звуков от их отражений. Задача создания модели (образа) акустического пространства - важная работа, выполняемая высшими отделами нервной системы.

Все эти свойства бинауральной слуховой системы (пространственная локализация, слияние слухового образа, эффект предшествования и др.), используются в настоящее время в развитии мощной индустрии "бинауральных технологий", включающих в себя создание программно-аппаратных средств, новых приборов, технологий звукозаписи и др. К числу наиболее эффективно развивающихся технологий в настоящее время можно отнести создание трехмерных виртуальных звуковых пространств (т.н. "аурализация", бинауральная стереофония, адаптивные процессоры и др.). Основные принципы их создания мы постараемся изложить в следующих публикациях.

Часть 6 Слуховая маскировка

Одним из самых важных свойств слуховой системы, широко используемым в современных технологиях цифровой звукозаписи, цифрового радиовещания и др., является эффект слуховой маскировки.

Желание передать по различным каналам звукового вещания все большее количество информации привело к разработке и широкому применению различных систем сжатия звукового сигнала (например, в стандартах MPEG), которые построены на использовании этого свойства слуховой системы.

Эффект маскировки связан с процессом взаимодействия сигналов, что приводит к изменению слуховой чувствительности к маскируемому сигналу (maskee) в присутствии маскирующего (masker).

Это взаимодействие тонов постоянно происходит в речи, где одиночные тоны практически не употребляются, и в музыке и приводит к тому, что восприятие сигнала в присутствии другого сигнала изменяется: меняется громкость, или сигнал вообще перестает быть слышимым (например, речь на фоне проходящего поезда), или изменяется восприятие каких-то отдельных спектральных признаков сигнала, то есть его тембр.

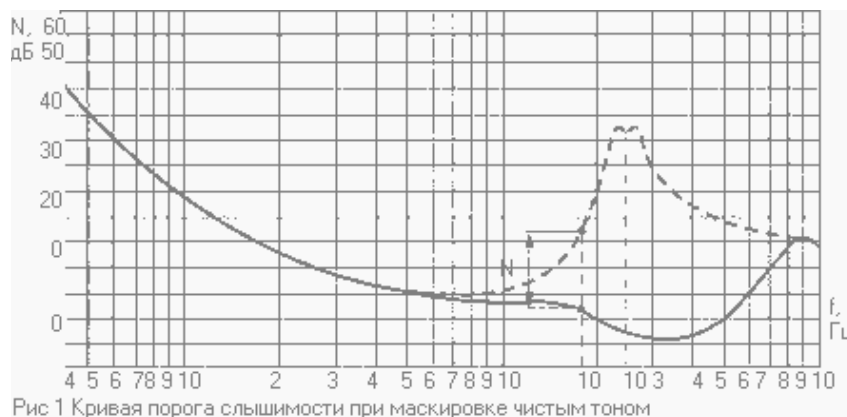
Процессы маскировки происходят в высших отделах головного мозга. Представим себе ситуацию: люди разговаривают, периферическая слуховая система принимает звуковые сигналы, обрабатывает и направляет в высшие отделы головного мозга, где они распознаются и оцениваются. Если в какой-то момент речи возникает сильный шум, то периферическая слуховая система продолжает принимать оба сигнала ? и речь и шум ? и направляет их в мозг. Однако в определенных отделах мозга речевые сигналы перестают восприниматься (не идентифицируются), и обрабатывается только шум.

Поэтому процессы слуховой маскировки ? достаточно сложное явление, и в настоящее время они находятся в стадии интенсивных исследований во многих мировых научных центрах, поскольку от их результатов в значительной степени зависит прогресс в современной цифровой звукотехнике.

Эффекты слуховой маскировки проявляются по-разному в зависимости от вида сигнала и способа его воздействия, и могут быть разделены на следующие основные группы:

- одновременное (моноуральное) маскирование;
- временное (неодновременное) маскирование;
- центральное (бинауральное) маскирование;
- бинауральное демаскирование;
- постстимульное утомление.

Остановимся на основных из них более подробно, поскольку в практической работе звукорежиссера умение пользоваться этими свойствами слуха играет очень большое значение при всех видах обработки звукового материала: при многодорожечной записи, монтаже, реставрации, введении различных эффектов и др.



1.Одновременное (моноуральное) маскирование звуков

Если в каждое ухо слушателя подавать основной тон с различной интенсивностью и частотой, то можно установить зависимость его порогов слышимости от частоты.

Если теперь к этому основному тону добавить дополнительный тон определенной частоты и интенсивности, то будет происходить взаимодействие обеих тонов, в результате произойдет изменение порогов чувствительности к основному тону (действительно, на фоне шума приходится сильно повышать голос, чтобы можно было его услышать). Если менять частоту основного тона, и на каждой частоте оценивать, на сколько дБ надо повысить уровень основного тона, чтобы можно было слышать его на фоне дополнительного мешающего (маскирующего) тона, то можно количественно оценить степень маскировки.

Степень маскировки есть разность в децибелах между уровнем порога слышимости данного тона в присутствии маскирующего тона и его уровнем порога слышимости в тишине.

На рис.1 показана кривая порогов слышимости основного тона в зависимости от частоты и повышение его порогов слышимости в присутствии маскирующего тона с частотой 2400 Гц и уровнем 60дБ. Из нее можно рассчитать степень маскировки основного тона на разных частотах: например, на частоте 1800 Гц степень маскировки dN ~15 дБ, на частоте 2000 Гц степень маскировки dN = 25 дБ и т.д.

Таким образом, количественно эффект маскировки оценивается по сдвигу (повышению) порога слышимости основного тона.

Измерения по указанной методике можно повторить для всех параметров основного тона (тестового стимула) и мешающего тона (маскера), и получить зависимости степени маскировки от частоты и интенсивности обоих тонов.

Анализ этих зависимостей позволил выявить интересные закономерности.

Маскировка, производимая определенным звуком, во многом зависит от его интенсивности и спектра. Еще в 1894 г. Мауег заметил, что, в то время как низкочастотные тоны эффективно маскируют звуки высокой частоты, высокочастотные тоны не обладают такими свойствами в отношении низких частот.

Маскировка, таким образом, является в отношении частот звука несимметричным эффектом.

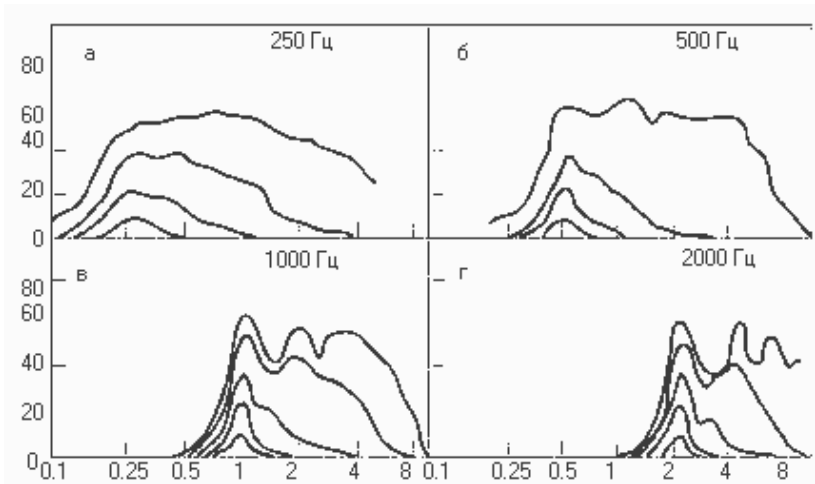


Рис 2 Зависимость степени маскировки от частоты и уровня маскирующего сигнала

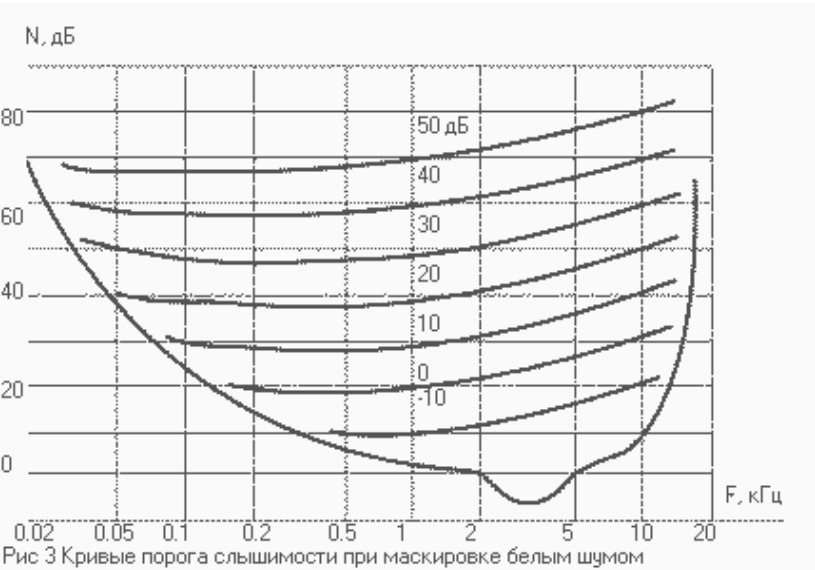


Рис 3 Кривые порога слышимости при маскировке белым шумом

На рис. 2 представлена серия образцов маскировки (иногда их называют аудиограммами маскировки). На каждом графике отражена степень маскировки, производимая определенным маскирующим звуком чистого тона, имеющим разную интенсивность. Здесь отложена только разность dN между уровнем порога слышимости в присутствии маскирующего тона и уровнем порога слышимости в тишине.

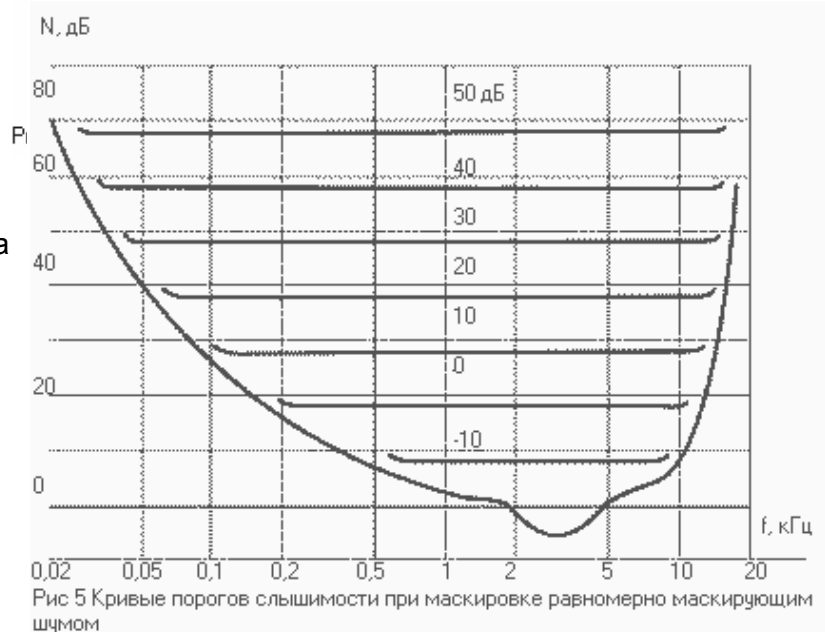
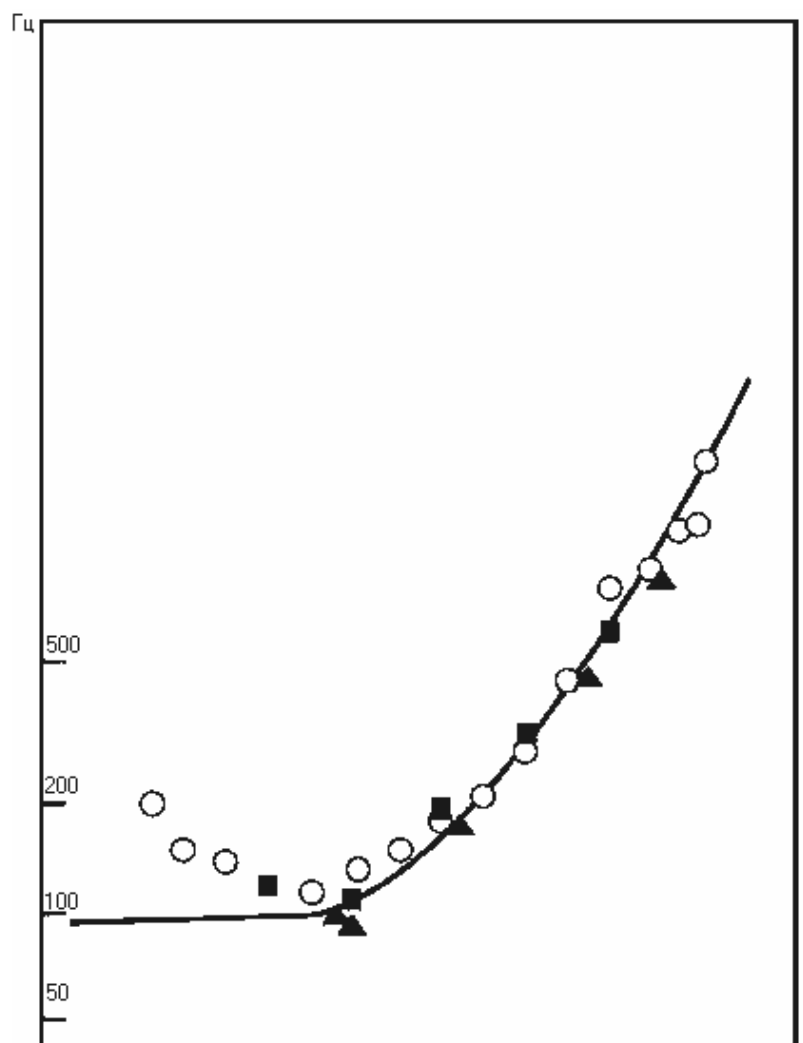
На рисунке 2 графически представлены следующие закономерности:

- а) наиболее выраженная маскировка наблюдается, если частота маскируемого звука близка к частоте маскирующего звука: степень маскировки уменьшается по мере увеличения разницы между той и другой частотой;
- б) степень маскировки увеличивается по мере нарастания интенсивности маскирующего звука (уровень его интенсивности в децибелах указан в виде цифр над кривыми);
- в) по мере нарастания интенсивности маскирующего звука маскировка становится все более несимметричной, выраженной по отношению к звукам высокой частоты;
- г) высокочастотные маскиры эффективно маскируют лишь звуки в относительно узком диапазоне частот, тогда как звуки низкой частоты являются эффективными маскерами для звуков в очень широком диапазоне частот.

Это явление связано со спецификой обработки звука в улитке уха (см. "Основы психоакустики", ч.1, "Звукорежиссер" № 6/1999). Максимум возбуждения низкочастотных звуков находится у самой вершины базилярной мембраны и низкочастотная огибающая бегущей волны имеет постепенно нарастающую амплитуду вдоль всей базилярной мембраны, достигая своего максимума у вершины с последующим крутым спадом. Таким образом, проходя вдоль всей базилярной мембраны, она оказывает воздействие и на ее нижние отделы, где находятся максимумы высокочастотных звуков. В то же время высокочастотные звуки анализируются только в нижней части мембраны, они не проходят к вершине и, следовательно, почти не оказывают влияния на низкочастотные звуки.

Таким образом, общее правило, что высокочастотные звуки маскируются сильнее, чем низкочастотные звуки, имеет очень большое значение в звукорежиссерской практике, например, при записи голоса и оркестра или нескольких разных инструментов со спектрами, расположенными в разных частотных областях и т.п. Всегда следует иметь в виду, что если уровень высокочастотных составляющих звукового сигнала будет недостаточно велик (его необходимую величину dN можно определить по кривым рис.2), то они будут замаскированы низкочастотными звуками.

Маскировка шумовыми сигналами - несмотря на то, что большую информацию об эффекте маскировки получают при исследованиях с тональными сигналами, при их использовании возникает ряд трудностей. Как известно, если два тона близки по частоте (разница меньше 15 Гц), между ними возникают биения. Кроме того, при больших интенсивностях могут отчетливо прослушиваться субъективные комбинационные гармоники, что затрудняет точные оценки эффектов маскировки. Поэтому были проведены исследования по количественному установлению степени маскировки, когда в качестве маскирующего сигнала выбирался узкополосный или широкополосный белый шум.



Результаты экспериментов узкополосным маскированием по существу подтверждают эффекты маскировки, наблюдаемые при исследованиях чистых тонов.

Результаты исследований эффектов маскировки при использовании широкополосного белого шума показаны на рис.3. Как видно из графика, степень маскировки зависит от уровня интенсивности шума маскера почти прямо пропорционально: увеличение интенсивности шума на 10 дБ вызывает увеличение порога слышимости (т.е. степени маскировки) тоже на 10 дБ.

Например, степень маскировки звука частотой 1000 Гц белым шумом со спектральным уровнем 40 дБ находят путем вычитания порога чувствительности к звуку с тоном в 1000 Гц в тишине (около 7 дБ) из порога восприятия этого же звука в присутствии шума со спектральным уровнем 40 дБ (приблизительно 58 дБ). Таким образом, степень маскировки составляет $58 - 7 = 51$ дБ. Если уровень шума составляет 50 дБ, то степень маскировки оказывается равной $68 - 7 = 61$ дБ.

Эти соотношения соблюдаются и для узкополосных шумовых и тональных сигналов, однако исследования последних лет показали, что на частотах ниже и выше маскера имеет место отклонение от линейного закона. Он соблюдается, когда частоты сигналов близки: для сигналов с частотой ниже маскера уровень маскировки растет медленнее, а для сигналов выше маскера - быстрее (увеличение на три децибела степени маскировки на каждый один децибел увеличения уровня маскера).

Кроме того, как следует из рисунка 3, белый шум неодинаково эффективен для маскировки разных частот: на низких частотах кривые (то есть степень маскировки) практически не зависят от частоты (примерно до 500 Гц). Но при дальнейшем увеличении частоты наблюдается четкая зависимость: при каждом удвоении частоты уровень порога слышимости повышается на 3 дБ. Причина этого заключается в наличии "критических полос слуха" (о которых было сказано в первой части цикла "Основы психоакустики").

Были поставлены многочисленные эксперименты, чтобы выяснить: вся ли ширина полосы белого шума участвует в маскировке данного тона, или существует определенная ограниченная (критическая) полоса, прилегающая к частоте тона, которая и дает в результате маскировку. Флетчер (Fletcher) показал, что если поддерживать спектральный уровень шума постоянным и расширять его полосу, то порог маскировки будет расширяться, однако как только полоса шума достигает определенной критической ширины, дальнейшее ее расширение не приводит к увеличению степени маскировки тона.

Таким образом, было показано, что только определенная "критическая" ширина полосы белого шума участвует в маскировке тона, равного центральной частоте этой полосы.

Наличие критических полос слуха отражает фундаментальную способность слуховой системы к частотному анализу, который выполняется во внутреннем ухе.

Максимальное смещение базилярной мембраны располагается в зависимости от частоты звука в разных ее местах. Можно считать, что на ней имеется линейка полосовых фильтров с определенной полосой пропускания. Ширине "слухового фильтра" соответствует расстояние примерно в 1...2 мм вдоль базилярной мембраны.

Если считать, что критические полосы примерно соответствуют ширине слуховых фильтров на разных частотах, то можно утверждать, что в маскировке участвует только та часть шума, которая попадает внутрь полосы пропускания фильтра с центральной частотой, соответствующей маскируемому тону.

Расширение полосы шума за пределы пропускания фильтра не увеличивает степень маскировки, несмотря на то, что громкость шума повышается.

В действительности не нужно представлять, что имеется серия дискретных критических полос, прилегающих друг к другу, - скорее следует представить, что каждая определенная частота сигнала расположена внутри полосы определенной ширины.

На рис. 4 можно видеть, что по мере увеличения центральной частоты ширина критической полосы

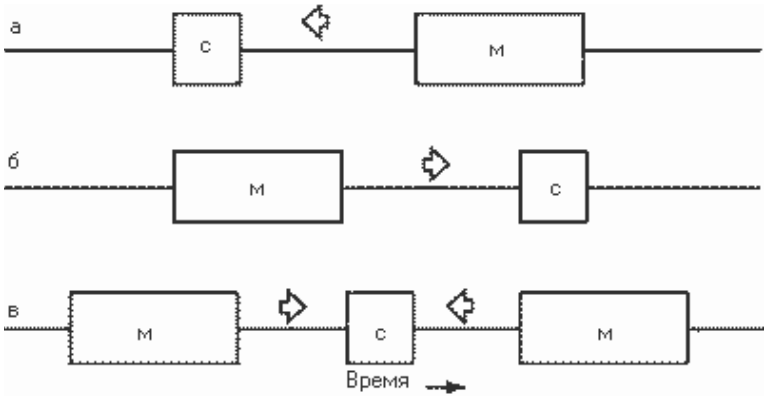


Рис 6 Последовательность включения маскера и маскирующего сигнала

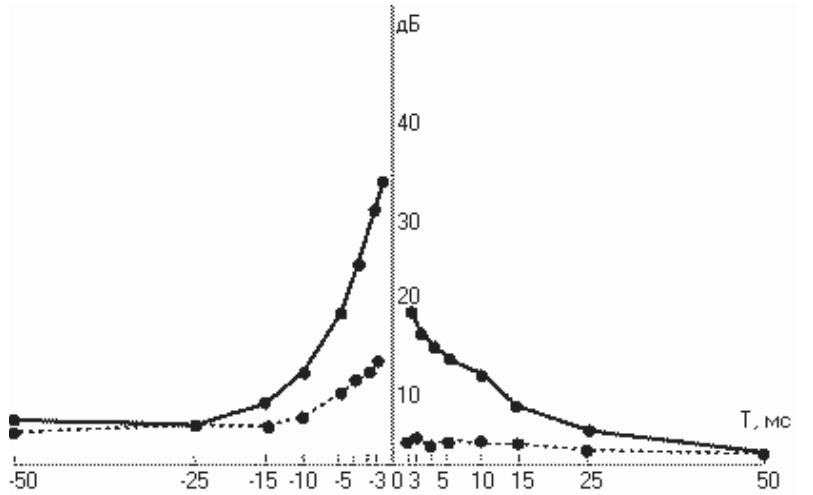


Рис 7 Значения степени маскировки (прямой и обратной) в зависимости от времени между сигналом и маскером

расширяется. Например, при центральной частоте 250 Гц ширина критической полосы равна 100 Гц, при центральных частотах 1000 и 4000 Гц соответственно 160 или 700 Гц.

Если вновь вернуться к рисунку 3, то можно объяснить, почему при удвоении частоты тонального сигнала степень его маскировки повышается на 3 дБ: это примерно соответствует закону пропорционального расширения ширины критических полос с увеличением средней частоты. Поскольку при этом расширяется полоса белого шума, участвующего в маскировке, то есть возрастает его общая интенсивность, то и степень маскировки соответственно увеличивается.

Можно подобрать шум с таким распределением спектральной плотности, чтобы он равномерно маскировал все частоты - для этого нужно, чтобы до частоты 500 Гц его спектральная плотность была равномерна и совпадала с белым шумом, а выше этой частоты спектральная плотность падала пропорционально частоте, как у розового шума.

Кривые порогов слышимости при маскировке равномерно маскирующим шумом показаны на рис.5.

Это свойство широкополосных шумов оказывать максимальное влияние на маскировку сигнала только в пределах критических полос положено в основу современных психоакустических алгоритмов сжатия сигналов в системах звукозаписи и радиовещания, где весь диапазон частот разбивается на ряд полос, примерно соответствующих критическим полосам слуха, и внутри каждой полосы производится расчет степени маскировки составляющих передаваемого сигнала.

2. Временное (неодновременное) маскирование

Выше порог маскировки был определен как "сдвиг порога восприятия одного звука, обусловленный присутствием другого звука". Все приведенные результаты рассматривали ситуацию, когда маскируемый и маскирующий (маскер) сигналы действуют на слуховую систему одновременно.

Однако в практике работы звукорежиссеров и музыкантов довольно часто возникают ситуации, когда достаточно громкие звуки маскируют, делают практически неслышимыми звуки, следующие за ними, а в некоторых случаях - даже предшествующие им. Такой вид маскировки, когда сигналы не перекрываются во времени, называется временной маскировкой.

Для изучения этого явления были поставлены эксперименты: (рис.6а, 6б, 6в)

Сигнал (например, удар, хлопок, импульс и т.п.) подается и выключается, а после короткой временной задержки подается маскер (другой достаточно интенсивный сигнал). Несмотря на то, что сигнал и маскер звучат не одновременно, возникает маскировка, то есть основной сигнал практически перестает быть слышимым. Такое соотношение получило название "обратной маскировки", поскольку подача исследуемого сигнала предшествует подаче маскира, т. е. эффект маскировки возникает в обратном направлении по времени (на рисунке показано стрелкой).

Как противоположность этому, существует предшествующая маскировка (рис. 6б). В этом случае маскер подается первым, а исследуемый сигнал включается через временной интервал после выключения маскира. Как указывает стрелка, маскировка сигнала предшествует во времени подаче маскира.

Степень маскировки исследуемого сигнала при подаче последующего (рис. 6а) или предшествующего (рис.6б) маскира определяется разными параметрами исследуемого сигнала и маскира. Этими параметрами являются:

- временной интервал между поступлением исследуемого сигнала и маскира * t ;
- уровень интенсивности маскира, дБ;
- длительность воздействия маскира * (мс) и др.

На рис.7 представлены некоторые результаты для оценки степени временной маскировки. На оси ординат отложены значения степени маскировки, вызываемой шумовыми сигналами (маскерами). Длительность воздействия маскира 50 мс при уровне звукового давления 70 дБ; частота исследуемого (маскируемого) чистого тона 1000 Гц, его длительность 10 мс. На оси абсцисс отложены временные интервалы * t (мс) между поступлением маскира и исследуемого сигнала при обратной и предшествующей маскировках. Наконец, сплошная линия обозначает степень маскировки при поступлении маскира и исследуемого сигнала в то же ухо (моноаурально), а пунктирная линия обозначает степень маскировки, когда маскер подается в одно ухо, а исследуемый сигнал - в другое (дихотическая маскировка).

На основании данных, представленных на рис.7, можно сделать следующие выводы:

- во-первых, обратная маскировка более эффективна, чем предшествующая. Другими словами, более высокий уровень степени маскировки наблюдается при поступлении маскира через короткий временной интервал вслед за сигналом по сравнению с маскировкой, выявляемой в том случае, когда исследуемый сигнал поступает через такой же интервал, но вслед за маскером.
- во-вторых, маскировка более выражена, когда сигнал и маскер подаются в одно ухо (моноаурально), чем тогда, когда исследуемый звук подают в одно ухо, а маскер - в другое (дихотически).

- в-третьих, сближение во времени подачи сигнала и маскира увеличивает маскировку. Наоборот, по мере увеличения временного разрыва между поступлением исследуемого сигнала и маскира степень маскировки уменьшается. Необходимо отметить, что степень маскировки резко падает при увеличении интервала от 0 до 15 мс, затем спад происходит плавно. Несмотря на то, что данные, представленные на рис.7, получены при временном разрыве "маскер/исследуемый сигнал" в 50 мс, была обнаружена значительная обратная маскировка для временных разрывов, превышающих 100 мс.

- в-четвертых, можно было ожидать, что временная маскировка будет увеличиваться по мере нарастания уровня интенсивности маскира. Однако для временной маскировки не найдено линейного повышения порога маскировки как функции уровня интенсивности маскира, характерного для описанной ранее одновременной маскировки. Таким образом, увеличение уровня интенсивности маскира на 10 дБ вызывает дополнительный сдвиг порога маскировки только приблизительно на 3 дБ.

- в-пятых, длительность действия маскира влияет на степень предшествующей маскировки, но не на обратную маскировку: так, маскер, длительность действия которого составляет 200 мс, вызывает большую маскировку, чем маскер, действующий в течение 25 мс.

- в-шестых, временная маскировка зависит от частотного взаимоотношения исследуемого сигнала и маскира точно так же, как и при одновременной маскировке. Другими словами, маскировка проявится в большей степени, если исследуемый сигнал и маскер весьма близки по частоте.

Интересно, что степень маскировки больше при сочетании обратной и предшествующей маскировок, чем суммарная степень маскировки при их раздельном исследовании. Сочетание маскировок осуществляется путем подачи исследуемого сигнала между двумя маскерами (см. рис. 6с).

Все это позволяет заключить, что обратная и предшествующая маскировки обусловлены разными механизмами.

О механизмах временной маскировки известно еще не достаточно: можно предположить, что при малых по времени интервалах поступления основного и маскирующего сигналов происходит взаимодействие (перекрывание) бегущих волн на базилярной мембране; при увеличении временных интервалов между ними до 200мс может сказываться инерционность нервных процессов в слуховой системе, например, - маскер, обрабатываемый нервной системой, подавляет процесс обработки исследуемого сигнала.

Все приведенные выше результаты по исследованию процессов временной слуховой маскировки могут быть полезны в практике работы звукорежиссеров, в частности при работе с электронными композициями, поскольку выбор последовательности звуков разной интенсивности с короткими временными интервалами между ними может привести к маскировке более тихих звуков (как предшествующих, так и последующих) более громкими (удар барабана, литавр, тарелок и др.), поэтому надо контролировать временной промежуток между такими сигналами и соотношением их интенсивностей.

Мы рассмотрели совпадающую и несовпадающую по времени маскировку сигналов, остальные эффекты ее проявления (бинауральная маскировка и демаскировка и др.) будут рассмотрены во второй части этой статьи.

Часть 7 Слуховая маскировка 2. Бинауральное маскирование

Как было отмечено в предыдущей статье - эффект маскировки связан с процессом взаимодействия сигналов, что приводит к изменению слуховой чувствительности к маскируемому сигналу (maskee) в присутствии маскирующего (masker).

Среди основных эффектов, связанных со слуховой маскировкой, можно выделить следующие:

- одновременное моноуральное маскирование;
- временное (неодновременное) маскирование (вперед и назад);
- центральное (бинауральное) маскирование;
- бинауральное демаскирование.

Первые два были уже рассмотрены, перейдем к анализу следующих.

Бинауральность слуховой системы обеспечивает не только локализацию в пространстве, повышение порогов чувствительности и др., но и позволяет получить интересные эффекты маскировки, которые вызывают сейчас очень большой научный интерес, поскольку позволяют судить о работе центрального нервного процессора, и могут найти большое прикладное применение.

Обычная маскировка происходит тогда, когда и маскируемый, и маскирующий сигналы поступают в одно и то же ухо, однако эффект маскировки возникает даже тогда, когда маскер и исследуемый сигнал подаются в разные уши. Этот процесс называется *центральным (или бинауральным) маскированием*.

Такое влияние маскира, вероятнее всего, обусловлено взаимодействием маскира и исследуемого сигнала на

уровне центральной нервной системы, где имеются специальные "бинауральные" нейроны, которые проводят сравнение сигналов от обеих ушей.

Центральное маскирование в некотором отношении подобно рассмотренному ранее маскированию при моноуральном слухе, хотя имеются и значительные отличия.

В целом, величина сдвига порога, вызванная центральным маскированием, гораздо меньше, чем при моноуральном маскировании, и проявляется в большей степени для звуков высокой частоты, чем низкой.

Степень маскирования становится значительной, только если время воздействия маскира не менее, чем 200 мс.

Особый интерес представляет частотная зависимость центрального маскирования. Наиболее выраженное маскирование выявляется, когда маскир и исследуемый тон близки по частоте. Частотная зависимость отражена на рис. 1, на котором маскир в виде тона 1000 Гц подают на уровне интенсивности 60 дБ в одно ухо, а маскируемый сигнал * в другое. Маскирование наиболее выражено в небольшом диапазоне частот, прилегающих к частоте маскира. Этот частотный диапазон совпадает с шириной критических полос слуха.

Рис. 1

Пик маскирования симметричен до 60 дБ (в отличие от моноурального маскирования), но при уровне выше 70 дБ уже появляется ассиметрия.

На рис. 1 показано также, что наибольшее маскирование происходит при пульсирующем маскире и исследуемом сигнале, (кривая 1), нежели при постоянно включенном маскире и пульсирующем в другом ухе сигнале (кривая 2). (Оба сигнала включаются и выключаются одновременно). Такие результаты получены при исследовании центральной маскировки, в разных экспериментах и у разных обследуемых.

Кроме того, по мере повышения уровня интенсивности маскира степень центральной маскировки нарастает только в случае подачи пульсирующего маскира и пульсирующего сигнала, тогда как при использовании постоянного маскира и пульсирующего сигнала степень маскировки сохраняется в пределах 1 и 2 дБ, независимо от уровня интенсивности маскира.

Поскольку реальная речь и музыка представляют собой постоянно изменяющиеся во времени (пульсирующие) процессы, можно предположить, что эффекты бинауральной маскировки особенно сильно оказывают свое влияние при прослушивании стереофонических и пространственных систем звуковоспроизведения, когда сигналы, поступающие на разные каналы слуховой системы, отличаются друг от друга.

Бинауральное демаскирование

Наибольший интерес в настоящее время вызывает эффект "бинауральной демаскировки", которому посвящены многочисленные статьи, доклады на конференциях, дипломы и диссертации.

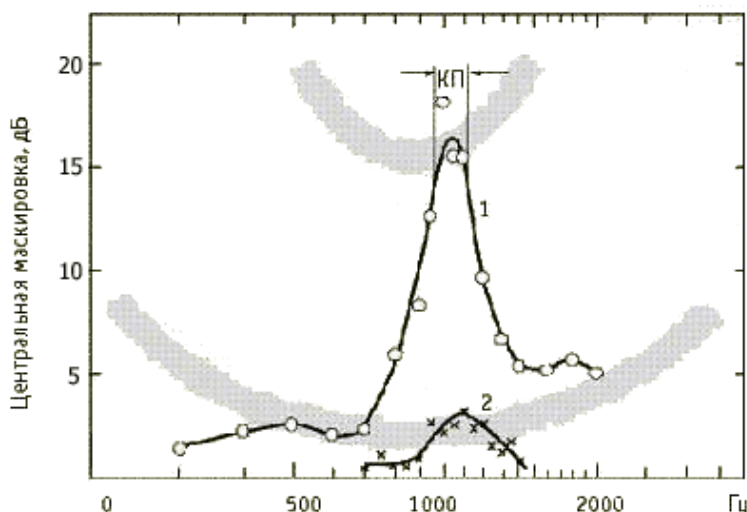
Эффект это проявляется в таком загадочном явлении: на фоне общего разговора (шума) можно "выслушать" интересующий слушателя разговор. Этот эффект получил название "эффект вечеринки" (Cocktail Party Effect).

Многочисленные исследования показали, что в основе этого явления лежит чувствительность к сдвигу фаз между сигналами при бинауральном слушании на частотах ниже 1500 Гц.

Бинауральное преимущество при маскировке значительно возрастает при стимуляции обеих ушей двумя различающимися стимулами. Такой способ подачи сигналов получил название дихотического. Результаты были исследованы как для тональных, так и для речевых сигналов.

Рассмотрим несколько возможных вариантов подачи сигнала и шума на два слуховых приемника (рис.2).

Представим типичный эксперимент по маскировке, в котором сигнал (С) маскируется шумом (Ш), причем уровень шума подобран таким образом, что он полностью маскирует полезный сигнал, например речь. Можно, пользуясь стереотелефонами, подать эту комбинацию на одно ухо СмШм (рис.2а), можно подать на оба уха СдШд (рис.2б) * в обоих случаях сигнал будет невозможно услышать на фоне шума. (Знак "м" означает моноуральную подачу, т.е. на одно ухо; знак "д" * дихотическую подачу, т.е. на два уха).



Если послать в одно ухо идентичный шум, а в другое * сигнал и шум (такой вариант СмШд представлен на рис. 2в), то тогда маскированный до этого сигнал вновь будет услышан (сигнал как бы освобождается от шума, его уровень субъективно повышается на 9 дБ).

Этот эффект и называется *бинауральной демаскировкой*.

При этом шум и сигнал локализируются в разных местах головы: шум * в середине головы, сигнал * ближе к одному уху. Получается, что шум и сигнал слышны в разных местах, и сигнал сразу обнаруживается из-за разной субъективной локализации.

По-видимому, что-то аналогичное происходит и на вечеринке: шум поступает с разных сторон, а нужный сигнал с одной стороны. Поворачивая голову, слушатель находит положение, при котором в ему оба уха поступает почти одинаковый шум. Тогда шумовой источник он слышит точно в центре, а сигнал локализуется в другом месте ближе к тому уху, через которое он поступает, поэтому сигнал начинает хорошо прослушиваться. Этот механизм срабатывает только при наличии в спектре низкочастотных составляющих.

В условиях экспериментальной ситуации (рис. 2а) слышимость маскированного сигнала можно еще увеличить путем изменения фазы шума в одном ухе на противоположную по отношению к шуму в другом ухе СпШд (рис. 2г) или изменения фазы сигнала в одном ухе на противоположную СдШп (рис. 2д). (Изменения фазы обозначено буквой *, поскольку стимулы различаются по фазе на 180°).

Изменение фазы на противоположную осуществляется путем изменения положительной или отрицательной полярности в одном из наушников.

Если определить "разность уровня маскировки" как различие (преимущество) в порогах маскировки при дихотическом прослушивании (т.е. при подаче разных сигналов на разные уши) и моноуральном (подаче на одно ухо), или при подаче на оба уха одинаковых сигналов, то величина этой разности количественно определяет уровень бинауральной демаскировки. Уровень бинауральной демаскировки показан на рис.2 (справа в виде численного значения в децибелах) для каждой комбинации сигнала и шума.

Величина разности уровня маскировки в зависимости от изменения параметров сигнала и шума колеблется от 0 до 15 дБ для СдШд, что представлено на рис. 2. Наибольшая разность уровня маскировки обнаружена при противоположных по фазе в обоих ушах сигнале (СпШд) или шуме (СдШп). Если сдвиг по фазе, например, для шума меньше π, то разность уровня маскировки уменьшается до 3...10 дБ (рис. 2е).

Как уже было сказано в первой статье по определению высоты тона (журнал "Звукорежиссер", 6/1999), разряды волокон слухового нерва связаны с фазой колебаний базилярной мембраны.

При этом на низких частотах звука степень привязки по фазе наибольшая, поэтому можно ожидать, что эффекты бинауральной демаскировки зависят от частоты сигнала.

Обобщенные результаты разных исследователей позволили установить, что если подать в оба уха шум одинаковый, а сигнал, по фазе разный (рис. 2д), то величина разности уровня маскировки, наибольшая для низких частот (около 15 дБ при 250 Гц), уменьшится по мере повышения частоты до величины 3 дБ (при 1,5...2 кГц).

Для объяснения эффекта бинауральной демаскировки была предложена модель работы слуховой системы, известная как модель Дурлаха, или "уравнивание - сокращение" (рис. 3).

В соответствии с этой моделью раздражение от звука, пройдя фильтры критических полос на базилярной мембране в каждом ухе (см. первую часть этой статьи), моноуральным и бинауральным путями поступит в слуховые центры мозга к "различающему устройству", которое определяет, присутствует ли полезный сигнал в данном шумовом окружении.

Различающее устройство переключается между тремя возможными каналами (два моноуральных и один бинауральный) и как основу для ответной реакции использует канал с наиболее подходящим соотношением сигнал/шум. Моноуральные каналы идут непосредственно к различающему устройству, а прохождение раздражения по бинауральному каналу включает две стадии: уравнивание и сокращение.

На первой стадии сигналы из обоих ушей уравниваются по амплитуде (уравнивающая стадия), затем в стадии

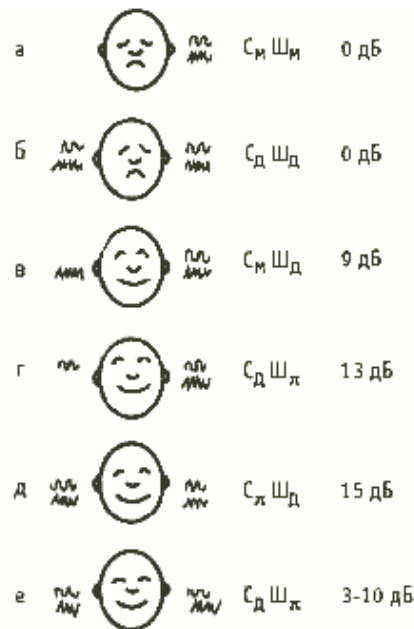


Рис. 2 Различия в уровнях маскировки при разных условиях подачи сигнала и шума



Рис.3 Структура модели "управление-сокращение"

сокращения эти сигналы вычитаются друг из друга. При одинаковых комбинациях сигналов в обоих ушах $C+Ш=C+Ш$ входной бинауральный сигнал полностью сокращается, и различающее устройство вынуждено выбирать между моноуральными каналами, поэтому разность уровня маскировки не выявится.

Однако при условии, что шум повернут по фазе, то есть когда $C+Ш=C+(-Ш)=2C$, происходит взаимное сокращение шумов и сигнал усиливается до 15 дБ.

Модель работает не безупречно из-за наличия внутриушного шума (поэтому выигрыш только на 15 дБ), точного поворота по фазе не происходит, не полностью уравниваются стимулы и т.д.

Разумеется, это только одна из возможных гипотез, исследования в этом направлении идут очень активно. Однако полученные результаты уже очень интересны, и могут породить новые неожиданные эффекты при многоканальной записи для пространственных систем звуковоспроизведения, построение которых требует учета бинауральных свойств слуха, в том числе и бинауральной демаскировки. Кроме того, использование бинауральной демаскировки может оказаться полезным для построения систем сжатия цифровых сигналов, что является одним из самых динамичных направлений в развитии аудиотехники.

Часть 8. Слуховые пороги, часть 1

Исследования способности слуховой системы воспринимать и преобразовывать в определенные слуховые ощущения (громкость, высоту, тембр и др.) основные объективные параметры звукового сигнала, такие, как интенсивность звука и пределы ее изменения (динамический диапазон), частотный диапазон, временные характеристики и т.д., является главной задачей современной психоакустики.

Слуховая система * чрезвычайно тонкий аппарат, но она имеет ограничения в восприятии частотного, динамического диапазона, в разрешающей способности, обладает нелинейными свойствами, очень чувствительна к перегрузкам и т. д.

Установление пределов возникновения слуховых ощущений, называемых слуховыми порогами, является в настоящее время одной из самых актуальных проблем в аудиотехнике, поскольку ее технические возможности значительно выросли за последние десятилетия, а возможности слуховой системы практически не изменились (а чувствительность даже несколько снизилась).

Современные системы звукозаписи и звуковоспроизведения, все звенья * от микрофона до громкоговорителя * проделали большой путь в деле усовершенствования параметров, и приблизились к порогам ощущений.

Идеология построения звуковой аппаратуры категории Hi-Fi (high-fidelity) состоит в обеспечении качества звучания, максимально близкого к живому звуку. Начиная с пятидесятых годов, времени появления Hi-Fi, в проектировании аудиоаппаратуры * акустических систем, микрофонов, усилителей и др. * произошли большие изменения в конструировании, технологии изготовления, компьютерном моделировании, технике измерений и т.д. Дальнейшее развитие звуковой техники зависит от успехов психоакустики, поскольку не имеет никакого смысла вкладывать средства в усовершенствование параметров (неравномерность АЧХ, нелинейные искажения и др.), если они уже

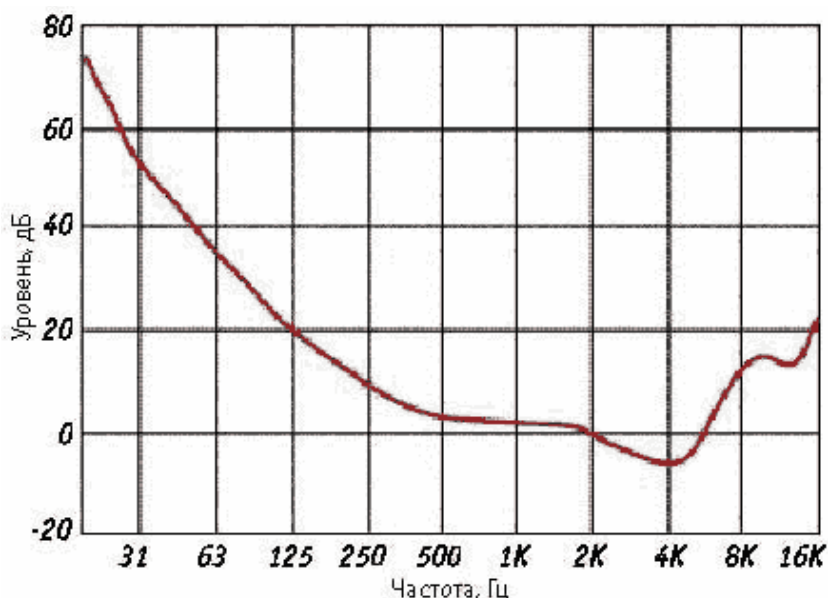
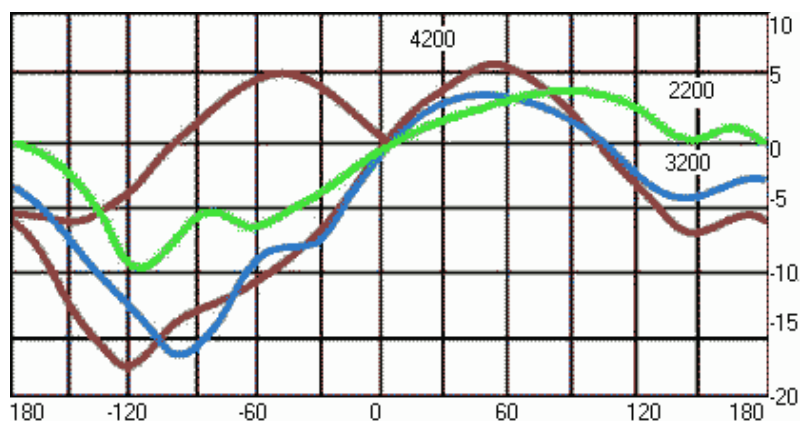
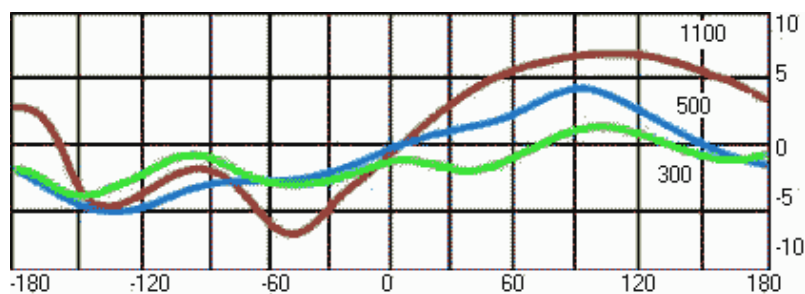


Рис 1 Кривая порогов абсолютной слуховой чувствительности, измеренная в условиях свободного поля



достигли порогов слышимости. Поэтому надо точнее определить пороги слышимости основных видов искажений, и направить усилия на поиск новых значимых для слуховой системы параметров. Точное определение слуховых порогов имеет принципиальное значение для современных систем цифровой обработки и передачи сигналов, в частности при выборе и построении систем сжатия и цифрового кодирования (стандарты MPEG и др.).

Широкое развитие систем пространственной передачи (Dolby Surround, бинауральной стереофонии и т.д.) также требует установления интенсивностных и частотных слуховых порогов, так как они определяют точность пространственной локализации.

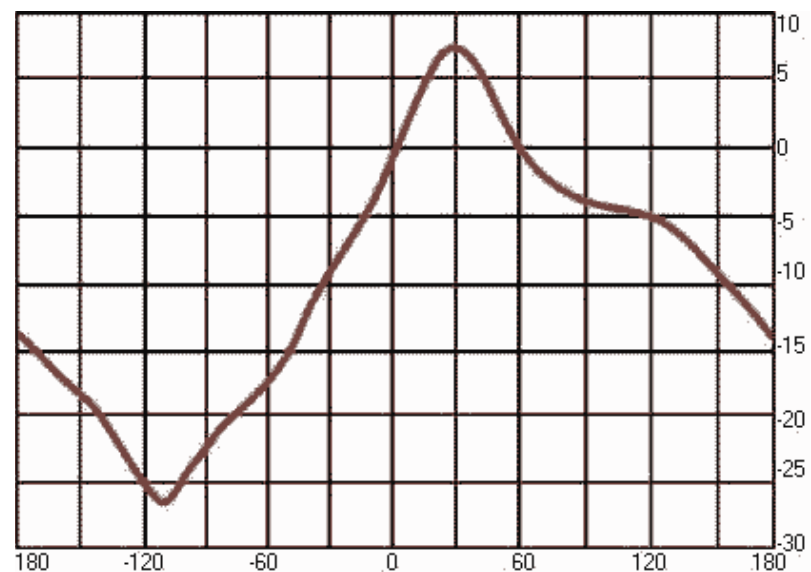
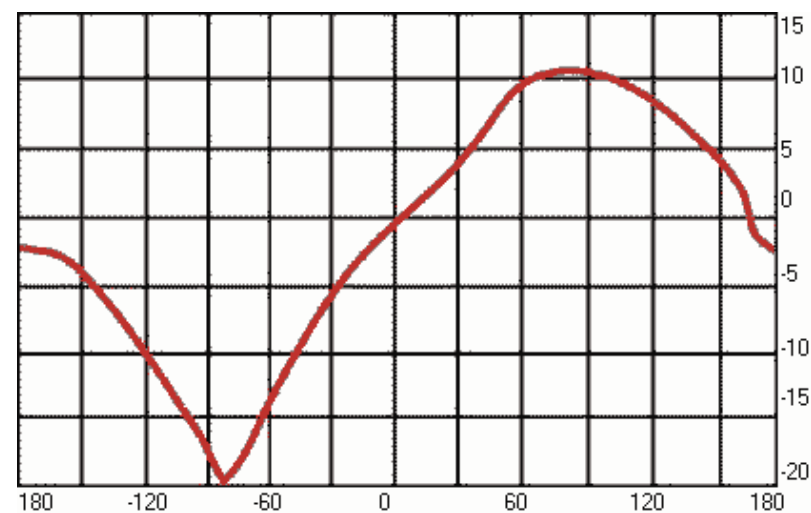
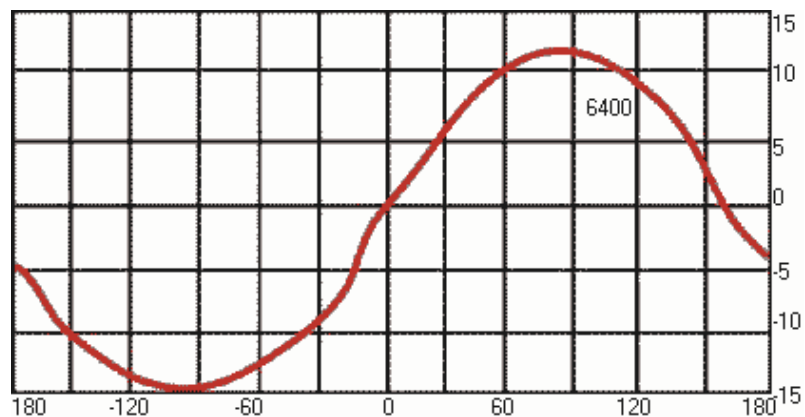
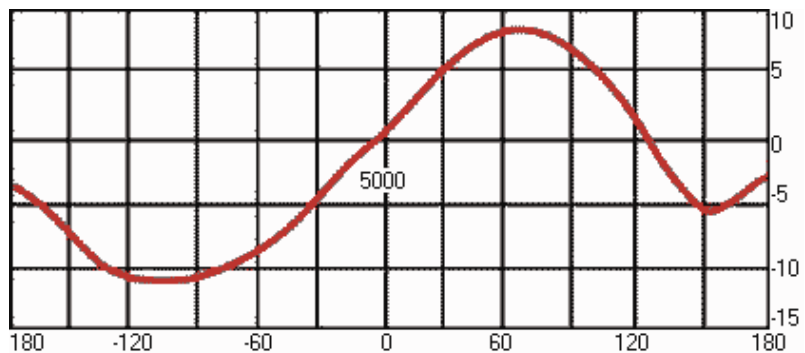
Все современные звуковые системы обработки и передачи музыкальных и речевых сигналов вносят определенные искажения в обрабатываемый сигнал. Главная задача их проектирования состоит в том, чтобы эти искажения были незаметны для слуха, т. е. лежали ниже порогов слуховой чувствительности.

Поэтому знание слуховых порогов имеет огромное значение для современной звукотехники, а, соответственно, и для работы звукорежиссера.

Слуховые пороги определяются минимальным значением объективного параметра звукового сигнала, при котором возникают слуховые ощущения. Они характеризуют чувствительность слухового аппарата к данному параметру * чем ниже слуховой порог, тем выше чувствительность.

Установленная в результате многолетних исследований картина слуховой чувствительности показывает огромные возможности слуховой системы:

- ухо человека улавливает звук, интенсивность которого 10-12Вт/м², т.е. 0 дБ (звуковое давление 2.10-5Па), с другой стороны оно ощущает как звук уровень давления 140 дБ - это соответствует отношению давлений 107 степени. Самый громкий звук в10 миллионов раз больше по звуковому давлению самого слабого;



- по частоте человек улавливает и очень низкие звуки, от 20 Гц и очень высокие, до 20 кГц (хотя музыкальные звуки в основном в диапазоне до 5000 Гц).

Необычайна чувствительность слуха к временным различиям (форме волны) и длительности звука. Чувствительность слуха к частоте, интенсивности и длительности связаны друг с другом. Слуховой аппарат имеет удивительную дифференциальную способность обнаруживать небольшие различия между сходными звуками по всем параметрам: интенсивности, частоте, временной структуре и длительности. Без этого невозможно было бы восприятие речи.

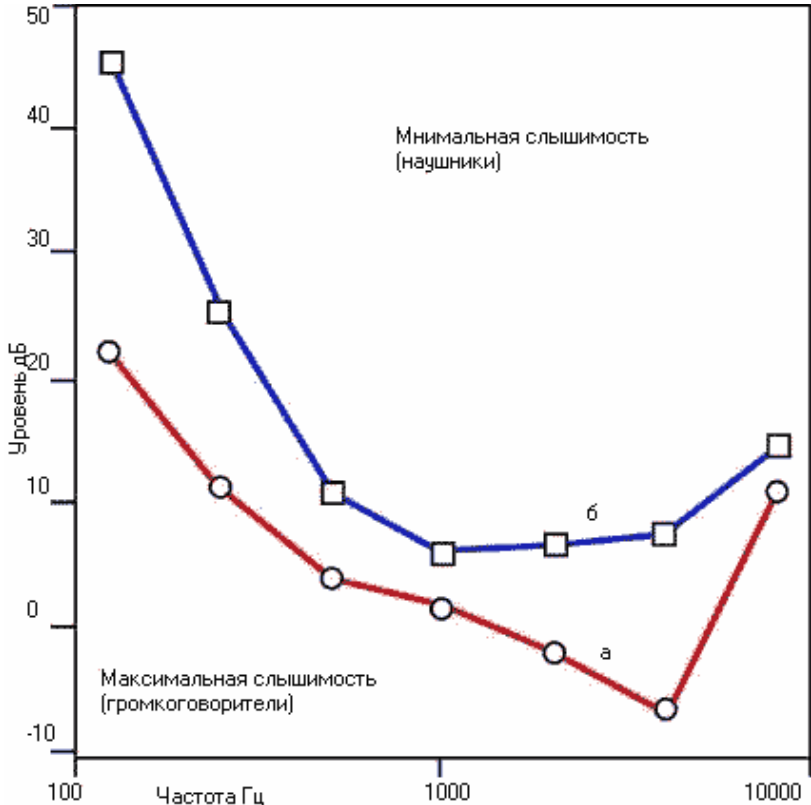


Рис 3 Кривые порогов слышимости а - при измерении в свободном поле б - через наушники

Абсолютные слуховые пороги

Не всякие изменения давления воспринимаются слухом как "звук" - существуют определенные границы слухового ощущения как по величине давления, так и по частоте. Эти ограничения зависят от уровня слуховых порогов.

Слуховые пороги могут быть разделены на абсолютные и дифференциальные.

Например, "абсолютный порог слышимости определяется как минимальное звуковое давление (в дБ), при котором еще возникает слуховое ощущение". Он характеризует чувствительность слуха к интенсивности звуковой энергии. Аналогично определяются абсолютные слуховые пороги по частоте и по временному интервалу.

Дифференциальные слуховые пороги характеризуют способность слуховой системы определять пороговое различие между звуковыми сигналами по частоте, по уровню звукового давления, по временному интервалу и др.

Опыты по определению абсолютного порога слышимости показали, что его величина зависит от условий опыта, особенностей звукового сигнала, параметров звукового источника и др.

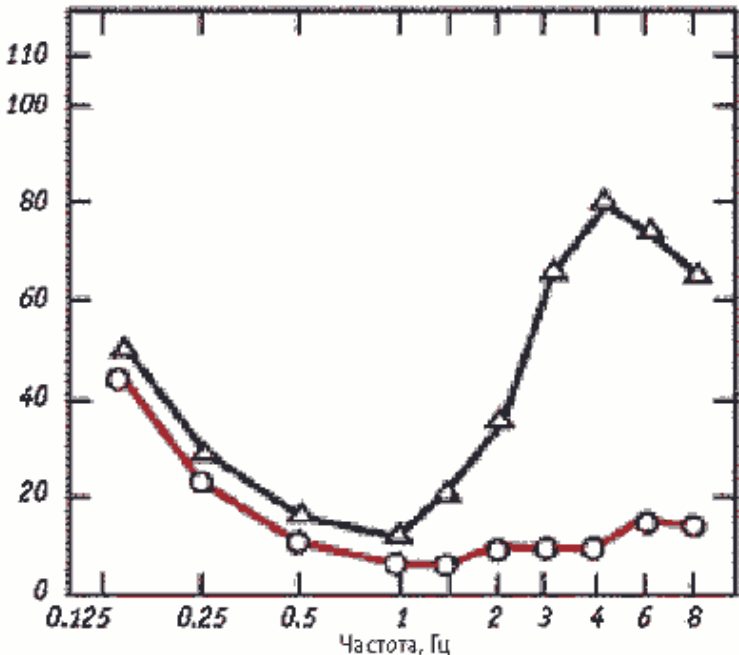


Рис 4а Аудиограммы, снятые относительно нулевого уровня

Под условиями опыта понимается характер звукового поля:

- создается ли оно одним громкоговорителем, помещенным перед слушателем, или многими источниками, равномерно распределенными вокруг головы;
- имеются ли отражения от границ помещения или приняты меры по их устранению;
- производились ли измерения минимального давления непосредственно около ушной раковины или в этой же точке при отсутствии слушателя (т. е. в свободном поле);
- предъявлялись ли сигналы через громкоговорители или через телефоны.

Обычно пользуются результатами измерений, полученными двумя основными методами.

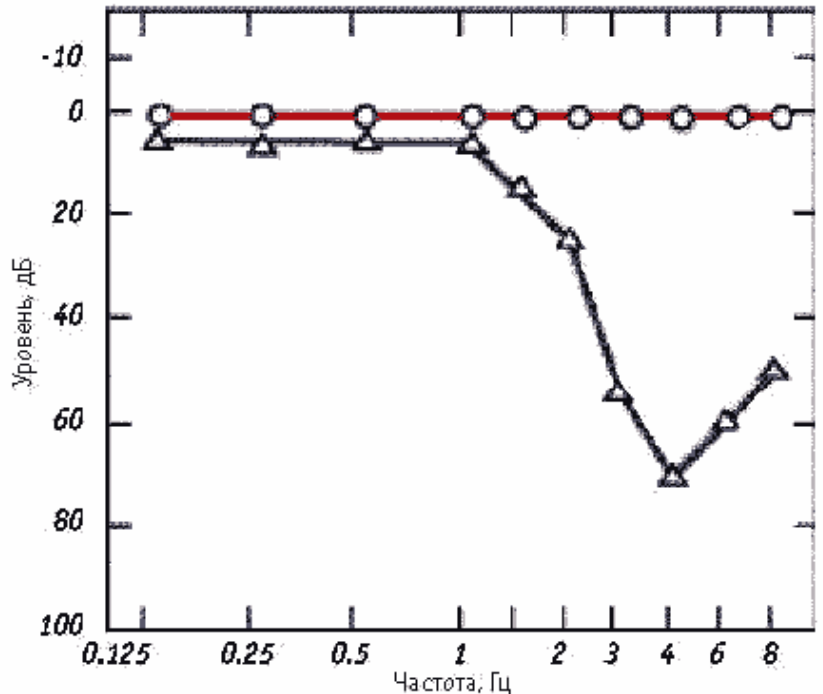


Рис 4б Аудиограммы, снятые относительно стандартных порогов

Первый: для свободного звукового поля (т.е. в заглушенной камере), формируемого одним излучателем, помещенным перед слушателем. Уровень звукового давления определяется микрофоном, помещенным в точку расположения головы слушателя. Измерение порога слышимости производится по методу так называемого балансного регулирования. Испытуемый имеет возможность с помощью переключателя менять направление изменения интенсивности звука, уменьшая его до уровня, когда тон становится неслышимым, и повышая до уровня, когда тон становится слышимым. Следовательно, регулируемый тон балансирует между значениями "слышен" и "не слышен".

Измерения проводятся на различных частотах, при этом на каждой частоте определяется полученный уровень. Описанные измерения должны быть проделаны с участием многих испытуемых, обладающих здоровым слухом, как тренированных (которые дают меньшие значения порогов), так и нетренированных.

Полученная таким способом кривая порога слышимости синусоидальных звуков, измеренная в условиях свободного поля, показана на рис. 1. Как видим, порог слышимости меняется в очень широких пределах в зависимости от частоты. Наибольшей чувствительностью ухо обладает в области частот 2500...3500 Гц, где порог слышимости имеет наименьшую величину. В области максимальной чувствительности слух воспринимает давление около 10-5 Па. Любопытно отметить в связи с этим, что звуковое давление, возникающее вследствие флуктуаций плотности воздуха, имеет при температуре 25°С величину 5×10^{-6} Па. Если бы ухо было вдвое чувствительней, оно слышало бы непрерывный шум флуктуаций молекул воздуха и тока крови. Таким образом, чувствительность слуха находится на пределе биологической целесообразности.

Таблица 1

Частота, Гц	100	200	400	800	1000	2000	3150	5000	8000	12500
Уровень, дБ	25,1	13,8	7,2	4,4	4,2	1,0	-3,6	-1,1	15,3	11,6

Следует отметить, что при бинауральном слушании слуховые пороги на 3 дБ ниже, чем при моноуральном.

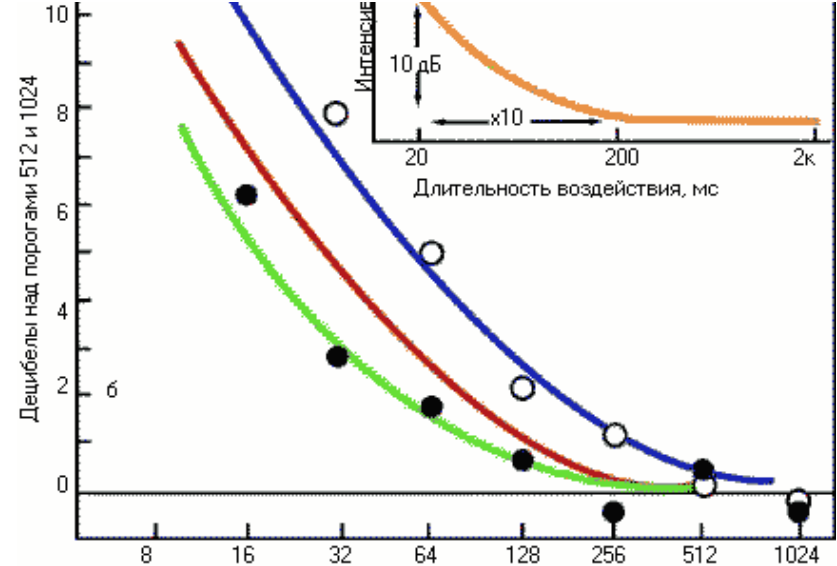


Рис 5 Зависимость порога слышимости от длительности сигнала при разных частотах

В международном стандарте ISO/R-226 (таблица 1) приняты за стандартные следующие значения порогов слышимости (громкоговоритель размещен в свободном поле на оси, слушатели в возрасте 18...30 лет):

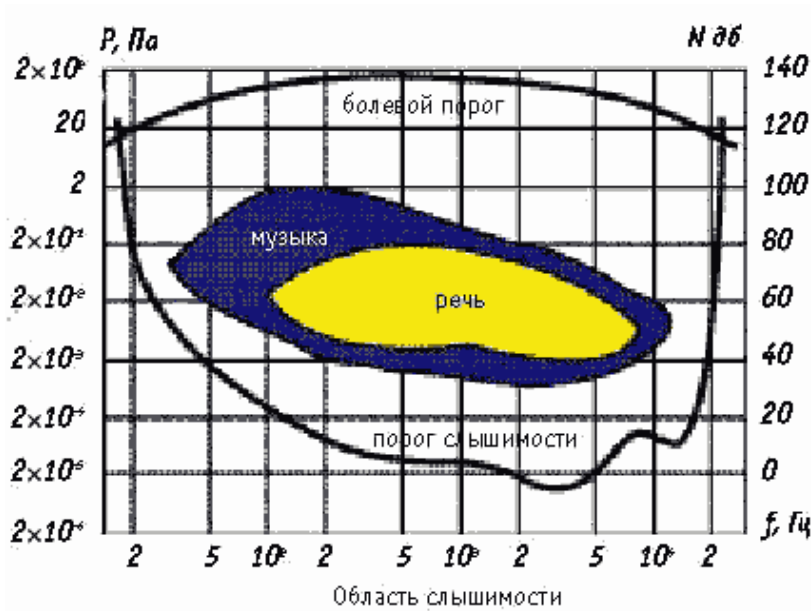


Рис 6 Кривые болевых порогов и область слышимости

Таблица 2

Частота, Гц	125	250	500	1000	2000	4000	8000
Уровень, дБ	45	25,5	11,5	7,0	9,0	9,5	13,0

Представленные на рисунке 1 и в таблице 1 значения абсолютных порогов слышимости относятся к случаю, когда источник-громкоговоритель размещен на оси 0о. Если перемещать громкоговорители под разными углами относительно головы слушателя, то абсолютные пороги существенно меняются. Значения порогов на различных частотах при разных углах размещения громкоговорителя показаны на рис.2. За нулевой уровень принято значение слухового порога, измеренного на оси 0о. Как следует из этих кривых, абсолютные пороги слышимости существенно меняются в зависимости от азимутального положения источника звука, и могут быть на некоторых частотах существенно ниже, чем на оси. Это объясняется фильтрующим влиянием ушной раковины и головы за счет дифракционных эффектов.

Таблица 3

SPL, дБ	90	92	95	97	100	102	105	110	115
T, часов в день	8	6	4	3	2	1,5	1	0,5	0,25

Второй метод * сигналы подаются на наушники, звуковое давление измеряется в слуховом проходе у барабанной перепонки. Кривые порогов слуховой чувствительности получаются несколько разными (рис.3): пороги при предъявлении сигнала через телефоны получаются выше, то есть слуховая система более чувствительна к сигналам, поступающим из внешнего пространства. Это объясняется тем, что за счет дифракции и резонансов ушной раковины, а также резонансов слухового канала, звуковое давление у барабанной перепонки усиливается в 2...3 раза, особенно в области частот 1500...3000 Гц, именно поэтому в области частот, совпадающей с резонансами слухового канала (~2700 Гц), и находится абсолютный порог слышимости слуховой системы.

При повышении и понижении частоты чувствительность слуха снижается и пороги слышимости соответственно повышаются.

В стандарте ANSI-89 для стандартных уровней слуховой чувствительности, используемых при калибровке аудиометров (приборов, измеряющих пороги слышимости), приняты следующие величины (при измерении на наушниках):

Абсолютные пороги слышимости существенно отличаются у индивидуальных слушателей в зависимости от возраста, состояния слуховой системы, наличия заболеваний и т. д. Для их оценки измеряются индивидуальные слуховые пороги и строится аудиограмма * график зависимости слуховой чувствительности от частоты. Она может быть построена как относительно абсолютных порогов, так и относительно нулевого уровня, за который приняты значения порогов из таблицы 2. Оба вида аудиограмм представлены на рисунках 4а и 4б.

Необходимо отметить, что всем, кто регулярно работает со звуком, в первую очередь музыкальным экспертам, звукорежиссерам и пр., необходимо регулярно снимать аудиограмму (для экспертов стандартом была установлена периодичность два раза в год).Для автоматического снятия аудиограмм в Интернете имеется специальная программа <http://www.digital-recording.com/audiomtri/audiomtr.htm> (можно попробовать).

Абсолютные пороги слышимости зависят от длительности предъявляемого сигнала: если длительность

сигнала мала (меньше 250 мс), пороги возрастают (рис. 5), и только при длительности больше 250 мс значения слуховых порогов стабилизируются к норме.

Если звуки имеют очень короткую длительность, то они воспринимаются как короткий щелчок. Требуется определенное время воздействия, чтобы можно было определить высоту тона, причем длительность этого отрезка времени зависит от частоты: при частоте 50 Гц требуется 60 мс, свыше 1000 Гц - 10 мс. Соответственно этому меняются, в зависимости от частоты, и уровни абсолютных порогов слышимости. Сокращение длительности воздействия с 200 до 20 мс на частоте 1000 Гц приводит к возрастанию порога на 10 дБ.

Это связано с особым свойством слуховой системы, называемым временной интеграцией (или суммацией). Слуховой аппарат работает как детектор энергии внутри определенного слухового окна длительностью примерно 200 мс. Требуется накопить определенное количество энергии внутри этого окна для достижения порога слышимости, причем чем короче сигнал, тем больше должна быть интенсивность звука, чтобы его можно было услышать, и наоборот. Ухо интегрирует энергию внутри этого временного окна, поэтому период времени 200 мс считается постоянной интегрирования слухового аппарата.

Болевой порог и область слышимости

Существует ограничение области слухового восприятия и со стороны громких звуков, хотя и не такое четкое, как порог слышимости. Например, синусоидальное звуковое давление с эффективным значением $p \sim 10$ Па (100 дБ) соответствует одному из порогов, называемому порогом неприятного ощущения. При достижении величиной p значения 60...80 Па (132 дБ) возникает ощущение давления на уши, подобное тому, которое бывает при закладывании ушей в самолете, а также неприятного щекотания в ухе. Эта величина называется порогом осязания. Наконец, давление 150...200 Па (140 дБ) причиняет боль и называется болевым порогом. Частотная зависимость болевого порога приведена на рис. 6.

Таким образом, динамический диапазон слуховой системы достигает 140 дБ, при этом акустическая мощность увеличивается в 45 раз. Существующая техника звукозаписи и звукопередачи, даже цифровая, еще не может обеспечить такую величину динамического диапазона сигнала. (Правда, имеются рекламные данные о микрофонах, способных обеспечить такой динамический диапазон, например микрофон 4138 фирмы B&K).

Нужно отметить, что слуховая система приспособлена к восприятию в основном тихих звуков и звуков средней интенсивности. Воздействие громких звуков (с уровнем выше 90 дБ) приводит к изменению порогов слуха и к необратимым изменением свойств слуховой системы, вплоть до полной глухоты. Причем степень повреждения пропорциональна времени воздействия громких звуков, поэтому международные стандарты (таблица 3) регламентируют допустимое время пребывания (Т час/день) в звуковой среде с высокими уровнями звукового давления, выше которых могут произойти необратимые изменения слуховой чувствительности:

Эта проблема особенно актуальна для звукорежиссеров, работающих достаточно длительное время с программным материалом с высокими уровнями звукового давления, а также для современной молодежи, испытывающей огромные перегрузки слухового аппарата на современных концертах и дискотеках, при прослушивании музыки на плеерах с ушными телефонами. Исследования, выполненные корпорацией BBC, показали, что уровни абсолютной слуховой чувствительности значительно снизились у молодежи за последние десятилетия.

Нередко после воздействия громких звуков высокой интенсивности у человека резко снижается слуховая чувствительность. Процесс восстановления обычных порогов может продолжаться до 16 часов.

Этот процесс называется "временный сдвиг порога слуховой чувствительности" или "постстимульное утомление". Сдвиг порога начинает появляться при уровне звукового давления выше 75 дБ и соответственно увеличивается при повышении уровня сигнала. Причем наибольшее влияние на сдвиг порога чувствительности оказывают высокочастотные составляющие сигнала. Величина сдвига порогов пропорциональна логарифму времени воздействия * поэтому и нормируется время прослушивания в день. Если измерять пороги чувствительности в разные сроки после выключения сигнала, то можно установить, что пороги начинают плавно снижаться, но примерно через две минуты происходит скачок в ходе восстановления чувствительности, а затем пороги продолжают плавно уменьшаться со скоростью, пропорциональной логарифму времени после выключения звука. Однако если время нахождения под воздействием громких звуков превышает допустимые нормы, то полного восстановления порогов чувствительности не происходит, постепенно чувствительность слуха снижается, что может привести к полной глухоте, особенно опасной, потому что она связана с повреждением волосковых клеток и поэтому практически не поддается лечению.

Абсолютные частотные пороги

Если на рисунке 6 посмотреть на кривые болевых порогов и кривые абсолютной слышимости, то можно видеть, что если продолжить эти кривые, то они как бы пересекаются, т.е. чтобы достичь порогов слышимости на самых низких и самых высоких частотах, требуются уже настолько высокие уровни, что они совпадают сразу с болевыми порогом, не создавая ощущения звука.

Таким образом, только звуки, попадающие в диапазон частот 20...20000 Гц, воспринимаются в виде слуховых

ощущений. Нужно отметить, что природа не наградила нас особенно острым слухом на высоких частотах, особенно если сравнить с собакой или кошкой, которые слышат до 60000 Гц, или дельфином (до 100000 Гц). Наверное, природа решила, что в этом нет никакой необходимости.

Измерения показали, что звуки с частотой 20 кГц могут услышать только очень редкие люди в очень молодом возрасте. В среднем чувствительность слуха к высоким частотам снижается каждые 10 лет на 1000 Гц. Примерно к 60 годам средний порог по высоким частотам составляет 12 кГц у женщин, у мужчин снижение частотных порогов происходит быстрее и часто составляет 5...6 кГц.

Однако если посмотреть на рис.6, то можно увидеть, что музыкальные и речевые сигналы занимают только часть слышимой области, как по частоте, так и по амплитуде. Основная энергия музыкальных звуков находится в частотной области от 40 до 5000 Гц, и по уровню звукового давления от 40 до 100 дБ, поэтому возрастное изменение частотных порогов приводит к некоторому уменьшению яркости звучания обертонов, но не мешает слушать музыку и речь, тем более что часто это дополняется большим музыкальным опытом.

Закончив на этом рассмотрение абсолютных порогов, во второй части статьи перейдем к рассмотрению дифференциальных порогов.

Часть 9. Слуховые пороги, часть 2

Ограниченные возможности слуховой системы определяются не только наличием абсолютных порогов слышимости, о которых было сказано в первой части статьи о слуховых порогах, но и ограниченной разрешающей способностью слуха.

Под разрешающей способностью слуха подразумеваются минимальные изменения звукового давления, частоты, временных интервалов (и соответствующих им громкости, высоты, длительности), которые могут быть замечены слухом. Разрешающую способность называют еще дифференциальным порогом восприятия (в англоязычной литературе JND just noticeable difference).

Современные компьютерные технологии открыли возможность вносить очень тонкие изменения в параметры звука, однако использование этих изменений должно опираться на знание разрешающей способности слуховой системы, иначе они останутся незамеченными (если только не ставится задача вносить какие-то специальные изменения, например, удалять короткие щелчки при реставрации и др.). Поэтому изучению дифференциальных порогов уделяется очень большое внимание и за последнее время получен ряд интересных результатов.

Амплитудные дифференциальные слуховые пороги

Вопрос о минимальных изменениях амплитуды давления, которые улавливаются нашим слухом, был исследован рядом авторов (Олсон, Цвиккер, Редерер и др.).

Постановка экспериментов по определению слышимых амплитудных различий сигналов у разных авторов различалась, однако полученные результаты позволили получить очень близкие значения JND.

Первая группа экспериментов использовала два синусоидальных сигнала одинаковой частоты, но разного уровня. Например, у входа в ушной канал подавался сигнал с частотой 1000 Гц с уровнем звукового давления 40 дБ, и второй сигнал той же частоты с изменяющимся уровнем. При поочередном прослушивании пары таких сигналов слушатель отмечал, какой из сигналов звучит громче. Естественно, что если разница в уровнях между сигналами достаточно большая (например, 40 и 60 дБ), то все 100% слушателей отметят эту разницу, но если разница в уровнях будет уменьшаться, то замечать разницу между сигналами будет все меньшее количество слушателей.

Разница в уровнях, которую замечает 75% слушателей, принимается за дифференциальный слуховой порог по уровню звукового давления (по амплитуде). Эти измерения, повторенные для разных частот и разных уровней звукового сигнала, позволили получить характеристики зависимости дифференциальных порогов слышимости JND от частоты и общей интенсивности звукового сигнала (рис.1). Как видно из рисунка, эти пороги (т.е. едва замечаемая разница в уровне громкости) зависят от частоты сигнала: наименьшие значения

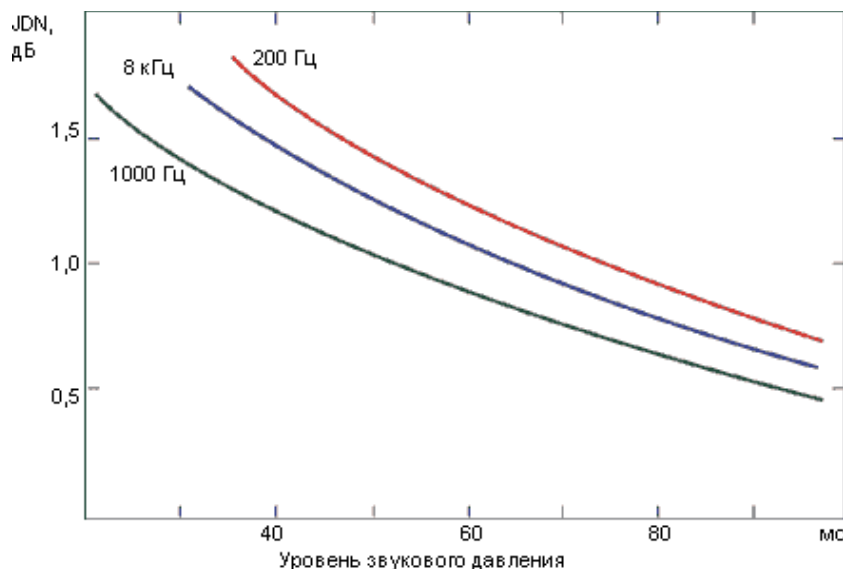


Рис 1 Зависимость дифференциальных амплитудных слуховых порогов от частоты и уровня сигнала

получаются на средних частотах(500...4000 Гц), на низких и высоких частотах они возрастают. Например, при общем уровне 60 дБ JND для частоты 1000 Гц составляет 0,8 дБ, а для частоты 200 Гц 1,3 дБ. Кроме того, они сильно зависят от общего уровня сигнала чем громче сигнал, тем меньшую разницу между сигналами можно услышать. JND на частоте 1000 Гц при общем уровне 40 дБ составляет 1,25 дБ, при уровне 80 дБ 0,6 дБ.

При другой постановке экспериментов использовался амплитудно-модулированный синусоидальный сигнал (пример такого сигнала показан на рис.2). Амплитудная модуляция сигнала достаточно широко используется в музыке (тремоло и амплитудное вибрато), она воспринимается на слух как небольшое изменение громкости сигнала.

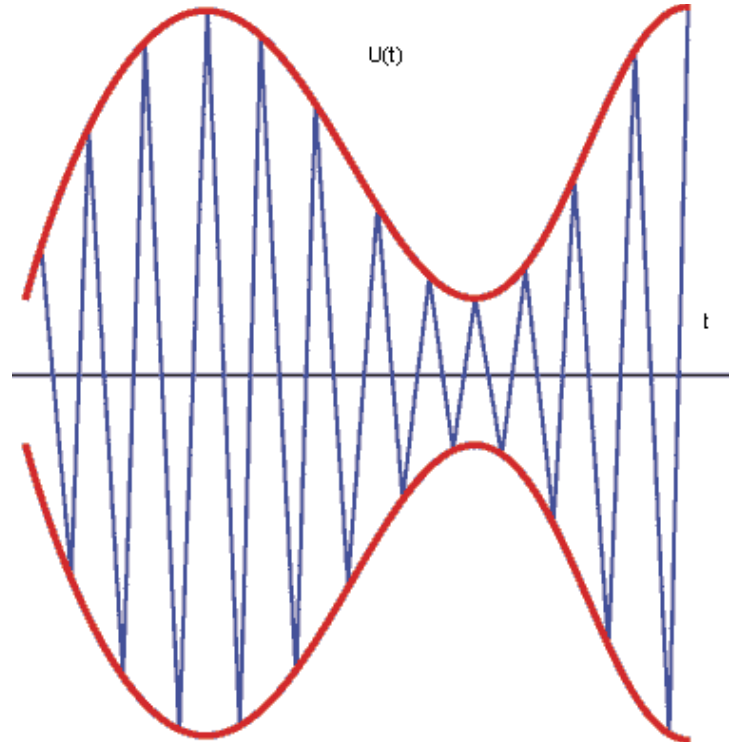


Рис 2 Амплитудно-модулированный синусоидальный сигнал

Предварительно был исследован вопрос о влиянии частоты модулирующего тона на заметность изменения амплитуды несущего сигнала. Наибольшая чувствительность слуха отмечена при частотах модуляции около 4 Гц, в связи с чем дальнейшие измерения производились при этой частоте.

Опыты сводились к определению уровня звукового давления, при котором становились заметными колебания громкости, обусловленные модуляцией. Результаты представлены на рис.3 в виде семейства кривых, которые можно назвать кривыми равной заметности амплитудной модуляции звука. Они почти повторяют рисунок кривой порога слышимости. Цифры, которыми обозначены кривые, выражают соответствующую каждой кривой глубину амплитудной модуляции в процентах. Из этих результатов также следует, что чем громче сигнал, тем меньшее изменение амплитуды модулирующего сигнала можно заметить. Например, при общем уровне сигнала около 90 дБ можно заметить изменение амплитуды всего 1,5%. При уровне сигнала 30...40 дБ, чтобы услышать изменение громкости, нужно менять амплитуду модулирующего сигнала приблизительно на 10% (это нужно иметь в виду при создании электронных композиций с введением эффектов модуляции).

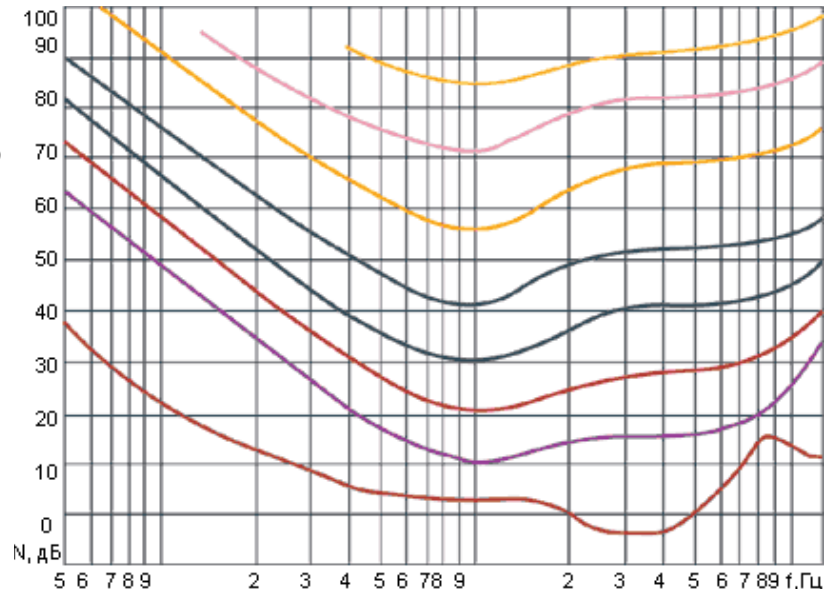


Рис 3 Заметность глубины амплитудной модуляции сигнала

На основе этих кривых можно получить зависимость амплитудной разрешающей способности слуха от частоты при постоянной громкости. На рис. 4 приведены такие кривые для уровней громкости 40, 60 и 80 фон (фон единица громкости, равная уровню звукового давления в дБ на частоте 1000 Гц, например, 40 фон = 40 дБ на 1000 Гц). По оси ординат отложено пороговое изменение звукового давления в процентах к уровню основного тона. Например, изменение $p/p_{100\%}=3\%$ при уровне 80 фон, т.е. 80 дБ (что соответствует значению $p=0,2$ Па), дает величину $p = 0,06$ Па.

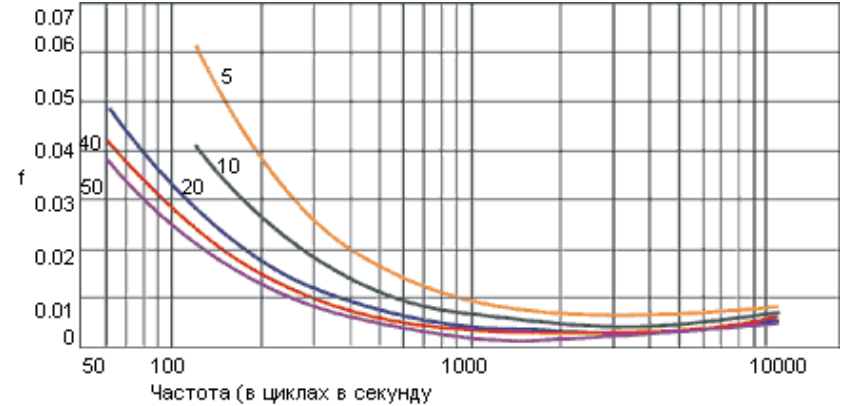


Рис 5 Зависимость дифференциальных частотных порогов от частоты

Амплитудная разрешающая способность слуха также сильно зависит от уровня громкости звука. Например, при частоте 1000 Гц для громких звуков (с уровнем громкости 80 фон) заметно изменение давления на 3%, в то время как колебания давления тихих звуков (40 фон) становятся заметными лишь при изменении на 10%. С

уменьшением громкости звука резче становится и частотная зависимость порога чувствительности слуха от изменения громкости. Таким образом, для чистых тонов с уровнем звукового давления, обычно используемым в музыке, замечаемая разница составляет от 0,5 дБ до 1 дБ в области средних частот.

Следует отметить, что для сложных музыкальных сигналов дифференциальные пороги существенно зависят от вида музыкальных программ (эстрадных, классических и др.), от опыта слушателя, свойств помещения и др.

Многочисленные эксперименты по определению чувствительности слуха к изменениям уровня звукового давления (т.е. неравномерности АЧХ) при воспроизведении через акустическую аппаратуру (громкоговорители, микрофоны и др.) показали, что пороговая величина воспринимаемых неравномерностей составляет в среднем 2 дБ, причем чувствительность слуха к пикам на АЧХ выше, чем к провалам. Кроме того, она зависит от ширины (добротности пика/провала) и его частотного расположения в области средних частот чувствительность максимальная.

Учитывая, что динамический диапазон слуховой системы около 120 дБ, то при такой тонкой чувствительности слуха к изменению уровней можно применять гораздо больше градаций по громкости (современные звуковые технологии позволяют это сделать), чем это используется в классической музыке, где указывается только шесть градаций от fff до ppp, что позволило бы существенно расширить таким образом средства музыкальной выразительности. (Классическая динамическая шкала предусматривает восемь динамических указаний: ppp-ppp-p-mp-mf-f-fff. В произведениях композиторов-романтиков Берлиоза, Вагнера, Чайковского встречаются и такие экстремальные обозначения, как rrrr и ffff, однако эти ремарки носят больше эмоциональный, чем реальный акустический смысл. Подробнее об этом см. 3/1999, стр. 30-31 - прим. ред.)

Частотные дифференциальные слуховые пороги

Частотная разрешающая способность слуха может быть определена путем прямых экспериментов: слушателю предъявляются два синусоидальных сигнала одинаковой интенсивности, и его просят менять частоту сигнала относительно опорного, пока он не услышит разницу по высоте. Эксперименты, выполненные для разных частот и разных уровней интенсивностей, позволили построить зависимости JND (дифференциальных порогов) от частоты.

Частотная разрешающая способность может быть также оценена по минимальным изменениям частоты, замечаемым слухом при частотной модуляции. Область повышенной чувствительности наблюдается при частотах модуляции около 4%.

На рис.5 приведена зависимость от частоты дифференциальных частотных порогов f (f/f 100%) для разных уровней громкости. Анализ кривых показывает, что пороги слышимости колебаний высоты тона зависят от частоты и интенсивности сигнала. В области частот до 1000 Гц при общем уровне звука 80 дБ f примерно составляет 3 Гц (в некоторых работах получены результаты до 1 Гц). Затем пороги начинают расти на 4000 Гц примерно до 10 Гц, выше определение различий по высоте резко уменьшается. Следует отметить, что слух замечает различие по высоте двух тонов при 3 Гц, но при этом при разнице частот до 15 Гц в суммарном звуке отчетливо слышны биения.

Как уже было отмечено в первой статье этого цикла (см. 6/1999), всего слуховая система различает 620 градаций высоты тона (140 градаций в диапазоне до 500 Гц и 480 градаций в диапазоне от 500 Гц до 16 кГц), что открывает большие возможности для развития микротоновой и спектральной музыки.

Временное различение звука

Способность слуховой системы различать тонкие временные различия в структуре сигнала является в настоящее время основным предметом многочисленных исследований. Причина этого заключается в том, что слуховой аппарат является принципиально нелинейной системой (как при больших, так и при малых уровнях сигнала), поэтому способность воспринимать различия параметров сигнала в частотной области не связана однозначно с восприятием временной структуры сигнала (механизм преобразования звука в слуховом аппарате не определяется преобразованием Фурье).

Именно этим можно объяснить тот факт, что акустические преобразователи (громкоговорители, микрофоны, акустические системы и др.), даже имеющие частотные искажения на уровне порогов слышимости, не обеспечивают качества звучания, идентичного с живым звуком (что, вообще говоря, и является главным критерием для аппаратуры Hi-Fi).

По-видимому, какие-то различия во временной структуре сигнала, которые до настоящего времени еще четко не определены, являются значимыми для слуховой системы, и именно по ним она определяет живое звучание или отличает одну акустическую систему от другой. Способность различать тонкую, быстро изменяющуюся временную структуру звукового сигнала подтверждается удивительно точным анализом и распознаванием речи, когда в непрерывном временном потоке распознается специфическая структура различных фонем.

Исследования разрешающей способности слуховой системы во временной области проводятся в нескольких направлениях:

Прежде всего, необходимо было выяснить, каково минимальное время, в течение которого ухо способно различать два сигнала. Это время можно измерить, предложив обследуемому различить два сигнала, одинаковые во всех отношениях, за исключением времени поступления. Это означает, что наибольшая чувствительность к временному различению является оценкой наикратчайшего периода времени, в котором слух способен интегрировать энергию звука. Можно оценивать этот период времени как низший предел шкалы временной интеграции.

В экспериментах по исследованию остроты слуха использовались щелчки или тональные импульсы. Оказалось, что характер задачи, поставленной перед обследуемым, крайне важен при измерении временных различий с помощью разных методов получают несколько различающиеся значения величин.

Допустим, что слушателю в быстрой последовательности подают два сигнала (высокий и низкий). Он способен ответить, поступает один или два следующих друг за другом сигнала, обнаружив разницу между началом сигналов в 2 мс. Эта величина не сильно зависит от частоты (временное различение даже обостряется для частот выше 1000 Гц), а также от интенсивности звука.

С другой стороны, ему необходимо время в 20 мс, чтобы определить, какой из сигналов поступает первым. Если необходимо оценить смысловое значение звука (речи, например), то это время увеличивается еще до 35 мс. Как уже было отмечено в первой статье, для определения высоты тона также требуется определенное время: для низких частот ~60 мс, для высоких ~15 мс.

Достаточно обученные слушатели способны идентифицировать быстрый ряд трехтональных раздражений, действующий в течение очень короткого периода 2...7 мс. При этом установлено, что на остроту временного слухового различения существенно влияют следующие факторы:

- число стимулов в ряду, каждый из которых должен быть обнаружен;
- способ подачи последовательных стимулов (раздельно или слитно);
- тип задачи, которую должен выполнить слушатель;
- степень его тренированности.

Следующей задачей было исследование дифференциальной чувствительности при воздействии звука с разницей в длительности dT .

Обследуемому предлагали два сигнала, один сигнал имел длительность T мс, а другой несколько большую длительность $(T+dT)$ мс. Интервалы поступали в случайном порядке, а обследуемый должен был указывать на

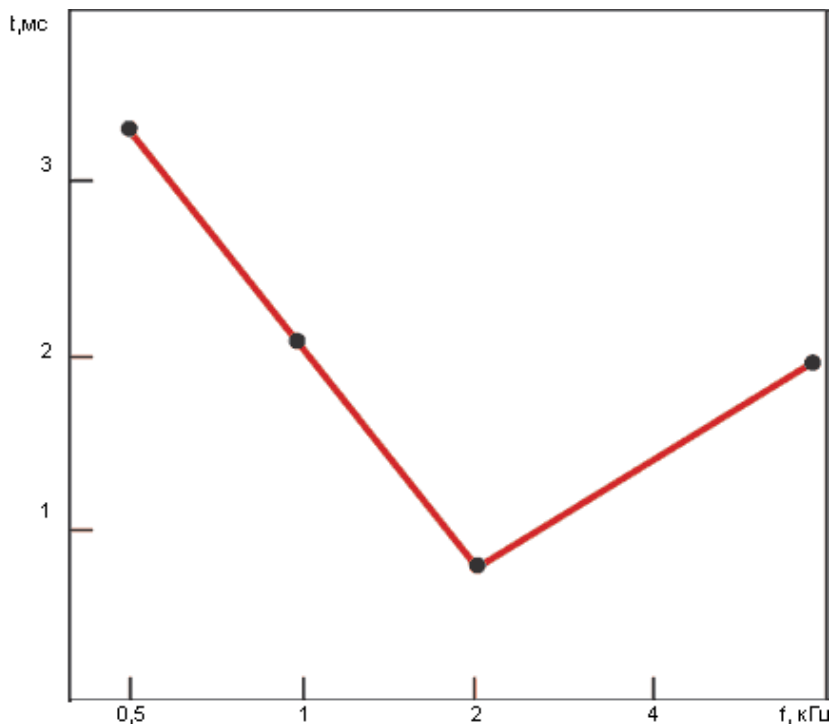


Рис 6 Пороги слуховой чувствительности ГВЗ

интервал с большей длительностью сигнала. Наименьшую разницу, правильно определяемую в 75% случаев, принимали за JND (дифференциальный порог) для длительности dT .

Основной вывод заключается в том, что dT начинает уменьшаться по мере уменьшения общей длительности воздействия сигнала (т.е. чем короче сигналы, тем меньшее различие по времени между ними слух может заметить).

Как можно видеть на рис. 6, dT уменьшается от 50 мс при длительности сигнала 960 мс, приблизительно до 0,5 мс при длительности сигнала меньше 0,5 мс. Дифференциальная чувствительность, dT/T (которая называется дробью Вебера) не представляет собой константу, а изменяется по мере изменения длительности так, что она равна $dT/T = 1$ при $T = 0,5 \dots 1$ мс, приблизительно 0,3 при $T = 10$ мс и 0,1 при $T = 50 \dots 500$ мс (T – длительность звукового сигнала). Результаты почти не зависят от ширины полосы и интенсивности звука.

Следующей важной проблемой было исследование чувствительности слуха к изменению времени установления (атаки) или спада сигнала. Время установления и спада звука является характерной особенностью различных музыкальных инструментов. Известно, что, меняя время атаки или спада сигнала, можно существенно изменить его тембр. У большинства музыкальных инструментов время атаки или спада лежит в пределах 5...360 мс.

Исследования дифференциальных порогов слуха для времени установления мс были выполнены для различных типов сигналов, и результаты, полученные для сигналов типа тональных посылок (прямоугольный импульс с синусоидальным заполнением) позволили установить, что дифференциальный порог для времени установления (как и для времени спада) для частот ниже 1000 Гц оказывается равным ≈ 1 мс, для частот $1 \dots 10$ кГц $\approx 0,5$ мс. Таким образом, изменения времени атаки и спада звукового сигнала, меньшие этих пределов, оказываются незаметными для слуха. Для реальных музыкальных сигналов эти пороги могут несколько отличаться в большую сторону за счет маскировки соседними звуками.

Разумеется, главной задачей современных исследований является установление слуховой чувствительности к тонкой временной структуре сигнала, в связи с чем особое внимание было уделено исследованиям дифференциальной слуховой чувствительности к фазовым искажениям. Изменения фазовых соотношений между спектральными составляющими сигнала существенно меняют его временную структуру. Однако на протяжении долгого времени, со времен Гельмгольца, считалось, что слух не чувствителен к фазовым соотношениям. Исследования последних лет показали, что это не соответствует действительности: изменения фазовых соотношений влияют на изменение тембра, четкость определения высоты музыкального сигнала и др.

В 80-е годы эти исследования привели к тому, что многие фирмы-производители Hi-Fi техники начали создавать аппаратуру с линейно-фазовыми характеристиками (в которых сигнал практически не претерпевал фазовых искажений), однако исследования Блаурта показали, что слух наиболее чувствителен к скорости изменения фазы, т.е. к групповому времени задержки (ГВЗ): $gr = -d(\varphi)/d\omega$.

В этих же исследованиях были установлены дифференциальные слуховые пороги для искажений ГВЗ (рис.7), которые для частоты 2000 Гц имеют минимальное значение ~ 1 мс. Эти данные используются в настоящее время при проектировании высококачественной акустической аппаратуры: искажения ГВЗ в них должны быть ниже установленных порогов.

Разумеется, полученные результаты не исчерпывают сложной проблемы установления порогов слуховой чувствительности к изменению временной структуры сигнала, и исследования в этом направлении продолжаются.

В заключение хотелось бы сказать о слуховой чувствительности к нелинейным искажениям, под которыми понимается появление в спектре звукового сигнала дополнительных спектральных составляющих. Это может явиться результатом компьютерной обработки или прохождения сигнала через электроакустический тракт.

Пороги слуховой чувствительности существенно зависят от характера нелинейности: при появлении низших (второй, третьей) гармоник пороги слуха для тональных сигналов составляют 0,1%, для фортепианной музыки 1...2%, для эстрадной музыки до 7%. Чувствительность слуха зависит от порядка гармоник: заметность гармонических искажений третьего порядка вдвое выше, чем искажений второго порядка, заметность искажений от пятого порядка и выше в 6...10 раз выше, чем второго. Именно этим объясняется странное явление, что в акустических системах, имеющих в основном нелинейные искажения низших порядков, пороговые значения составляют 1...2%, в то же время в транзисторных усилителях и цифровой аппаратуре, где возникают нелинейные искажения высоких порядков, уровни нелинейных искажений должны составлять сотые и тысячные процента, чтобы они были незаметны для слуховой системы.

Как уже было отмечено выше, современные компьютерные технологии открывают очень широкие возможности при обработке звука, однако при всех видах обработки следует учитывать возможности слуховой системы, для чего и необходимы данные как по абсолютным, так и по дифференциальным слуховым порогам.

Часть 11. Аурализация - виртуальный звуковой мир

В статье "Научные результаты 108 конвенции AES" ("Звукорежиссер" №3/2000) мною было обещано сделать три вещи:

- передать CD-ROM с докладами конвенции в редакцию (что было выполнено);
- подробнее рассказать о принципиально новом направлении в создании пространственных звуковых полей компьютерной технологии аурализации (об этом в данной статье);
- познакомить с направлениями исследований в мировом центре компьютерной музыки и акустики IRCAM (а об этом в следующих номерах).

Итак, начнем с самого "горячего" направления в современной звукотехнике - "аурализации".

В одной из статей фирмы "Брюэль и Кьер" была высказана любопытная мысль, что отношения человека со звуком можно разбить на три крупных этапа:

I. От начала эпохи "человека разумного" до начала XX когда звук прошел эволюцию от средства чисто сигнального, необходимого для выживания, через средство коммуникации (речь) к средству эмоционального и эстетического воздействия, то есть к музыке.

К началу XX века музыка достигла невиданных высот, стала мощным средством передачи величайших глубин человеческой мысли и эмоций. Гениальные композиторы (Бах, Бетховен, Моцарт и др.) подняли музыкальное творчество на необыкновенную высоту, разработали особый язык (код), способный передавать не меньшее богатство мыслей и нюансов, чем письменная и устная речь (литература).

Человечество создало широкую палитру музыкальных инструментов, отработало вокальную технику, построило великолепные концертные залы, театры, соборы и др. Однако, эти величайшие достижения искусства были доступны очень ограниченному кругу людей лишь доли процента населения могли слушать хорошую "естественную" музыку в хороших "естественных" залах);

II. от начала XX века до 80-х годов с момента изобретения радио и телевидения музыкальное и вокальное искусство стало доступно миллионам, но, как всегда, при массовом тиражировании качество звука резко упало отставали технические средства. Главной задачей в тот период было передача смысловой (семантической) вербальной информации.

К середине века техника звукозаписи, воспроизведения и звукопередачи значительно выросла, и это позволило поднять проблему передачи эмоциональной и эстетической информации на новый уровень родилось движение Hi-Fi (high-fidelity, высокая верность воспроизведения), идеология которого состояла в том, чтобы акустическая аппаратура могла воспроизводить звук максимально похоже на натуральный "живой" источник. Постановка проблемы достоверной передачи звука в записи дала мощный толчок к развитию акустики и созданию крупной индустрии, производящей звукотехническую аппаратуру.

Современная акустика представляет мощное и развитое направление науки во всех странах мира, и имеет огромную промышленную базу: сотни научных институтов, тысячи фирм, разрабатывающих и производящих огромное разнообразие звукотехники:

- студийное оборудование микшеры, микрофоны, усилители, мониторы и т.д.;
- передающее оборудование радио- и телепередатчики;
- воспроизводящее оборудование акустические системы, проигрыватели, магнитофоны и др.

Объемы мирового производства например, громкоговорителей, достигают более 500 млн в год; акустические системы только на рынке США представлены 300 фирмами, выпускающими более 3000 моделей) и т.д.

III. от 80 лет до настоящего времени. В начале 80-х известным акустиком проф.Олсоном (Olson) были сформулированы проблемы третьего этапа развития акустики "перенос атмосферы первичного поля в любое вторичное помещение прослушивания"

Сама постановка такой проблемы была бы в принципе невозможна без создания новой научной и технической базы развития цифровых компьютерных технологий обработки звука и соответствующих технических средств: звуковых процессоров, цифровых станций обработки, монтажа, редактирования, архивирования и т.д., цифровых магнитофонов и лазерных проигрывателей и т.д.

Разработка этой проблемы уже привела к развитию пространственных систем звукопередачи (Dolby, бинауральная стереофония и т.д.), появлению адаптивных цифровых процессоров, систем пространственного звуковоспроизведения Dolby Stereo, Dolby surround и др. Однако она потребовала решения целого ряда новых задач в области психоакустики, и привела к появлению новых направлений в цифровой акустике.

Одним из принципиальных этапов в решении этой проблемы было создание технологии аурализации.

Термин "аурализация" (auralization) появился несколько лет тому назад и еще не определился окончательно. Его определение дал Мендель Клейнер (Mendel Kleiner) по аналогии с термином "визуализация" на конгрессе AES в 1989 году. Он звучит так:

"Аурализация процесс превращения звукового поля источника в пространстве в "слышимый звук" путем физического или математического моделирования таким образом, чтобы смоделировать бинауральное слуховое ощущение на заданной позиции моделируемого пространства".

Сейчас этой проблеме уделяется очень большое внимание в специальной литературе: появляется много научных статей и докладов, в том числе и на последнем конгрессе AES; разработаны специальные пакеты программ для реализации этой идеи; появились фирмы, которые специализируются на создании и внедрении таких программных продуктов, например, фирма Одеон, которая предложила свое определение:

"Аурализация искусство создания цифровых моделей бинауральных записей в несуществующих помещениях".

Иначе говоря, *аурализация - это способ воссоздать трехмерное звуковое поле, пытаясь с помощью компьютерных программ повторить способы обработки звука, которые слуховая система применяет к звуковому сигналу в помещении, чтобы создать ощущение пространства.*

Необходимо отметить, что точное определение этого процесса пока еще не принято окончательно, а в русской технической литературе его вообще еще нет.

Нужно сказать, что трехмерное визуальное пространство удалось создать раньше, что нашло уже широкое применение в компьютерных играх, видеоклипах, системах обнаружения и др., поэтому моделирование трехмерного звукового пространства стало необходимым этапом, поскольку вместе они могут полностью воссоздать "пространственный виртуальный мир". К чему это приведет в XXI веке сказать трудно...

Попробуем рассмотреть, что надо сделать с музыкальным сигналом, чтобы после его компьютерной обработки слушатель, находящийся при воспроизведении в любом помещении, воспринимал звуковое пространство таким же, как если бы он слушал музыку в реальном концертном зале.

По существу, задача ставится таким образом: как надо "обмануть" мозг, чтобы создать у слушателя ощущение трехмерного звукового пространства концертного зала вот для чего нужна психоакустика. Рис.1.

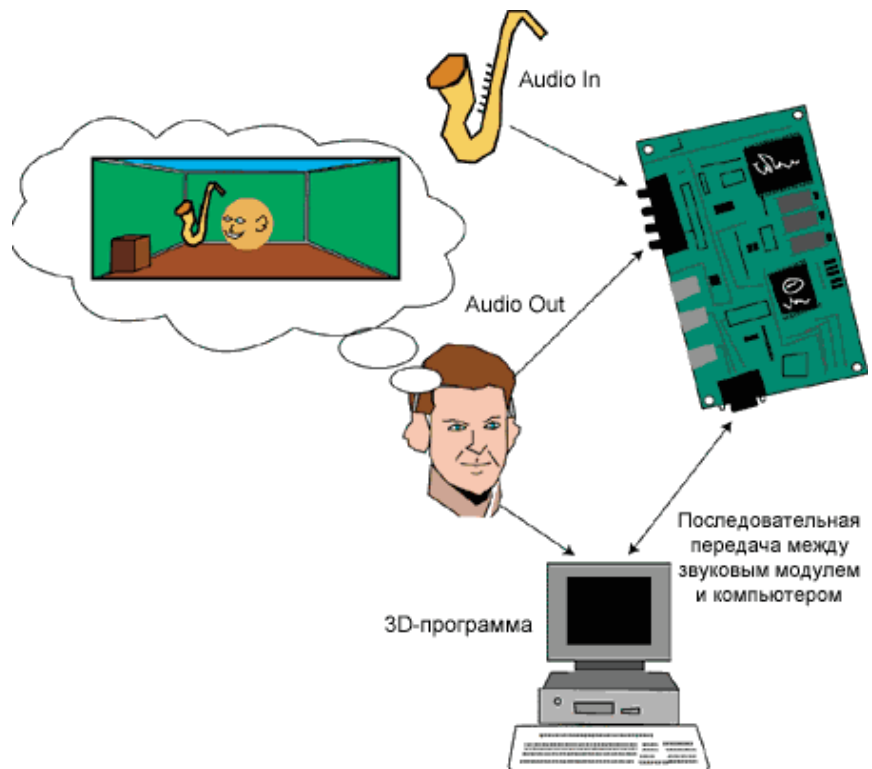
Рис.1. Общая структура системы аурализации

Как уже было рассмотрено в предыдущих статьях по психоакустике (и еще будет рассмотрено в следующих), только два физических параметра сигнала воспринимаются нашей слуховой системой: интенсивность (т.е. энергия или звуковое давление), и время начало и конец сигнала, и его повторяемость во времени (периодичность или частота).

Человек "слышит" звук, воспринимая изменения величины звукового давления, воздействующие на его барабанную перепонку, во времени. Вся информация, которую мы получаем о звуке, содержится в звуковых волнах, являющихся сжатием-разрежением воздуха. Все остальное, что мы оцениваем в звуке: его громкость, высота, тембр, звуковое пространство, тонкие музыкальные нюансы и др. - это результат обработки его нашим слуховым аппаратом и мозгом.

Рассмотрим, что происходит, когда слушатель воспринимает звук в помещении.

Музыкальный инструмент (голос, оркестр и т.п.) создают определенный акустический сигнал, который представляет собой некоторую звуковую волну, с определенной зависимостью звукового давления от времени $p_1(t)$. Например, осциллограмма звучания мужского голоса при произнесении слова "sound" (записанная в заглушенной камере) показана на рис.2. Затем этот сигнал определенным образом изменяется помещением за счет отражений звуковых волн, процессов затухания, дифракции и т.д. Если рассматривать помещение как линейный фильтр, который имеет свои характеристики $p_{\text{пом}}(t)$, то в каждой точке пространства суммарный сигнал получается как "свертка" сигнала источника и характеристик помещения (термином "свертка" называется результат обработки одного сигнала другим, например, в данном случае



$$p_{\Sigma}(\tau) = p_1 \oplus p_{\text{пом}} = \int_{-\infty}^{\infty} p_1(\tau) p_{\text{пом}}(\tau-t) dt.$$

Рис.2. Спектрограмма звука sound, записанная в заглушенной камере

Наконец, голова и ушные раковины производят свою обработку звукового сигнала (см. "Звукорежиссер", №10/1999). Пример осциллограммы того же самого звука после обработки его помещением и слуховой системой в левом и правом ушах показан на рис.3.

Рис.3. Спектрограмма звука sound, снятая на двух входах в ушной канал в помещении

Таким образом, чтобы вызвать такие же ощущения у слушателя, надо сделать следующее:

записать оригинальный звуковой источник, изменить программно его импульсную характеристику (или спектр) так, как это делает заданное помещение (концертный зал, собор или подвал), а затем произвести его фильтрацию аналогично тому, как это происходит в ушных раковинах и голове. После того как это сделано, можно воспроизвести эти сигналы, например, через головные телефоны, и получить ощущение того, что слушатель находится в реальном трехмерном акустическом пространстве вместе с источником звука (певцом, музыкантом, оркестром) см. рис.4 .

Рис.4. Схема обработки сигнала

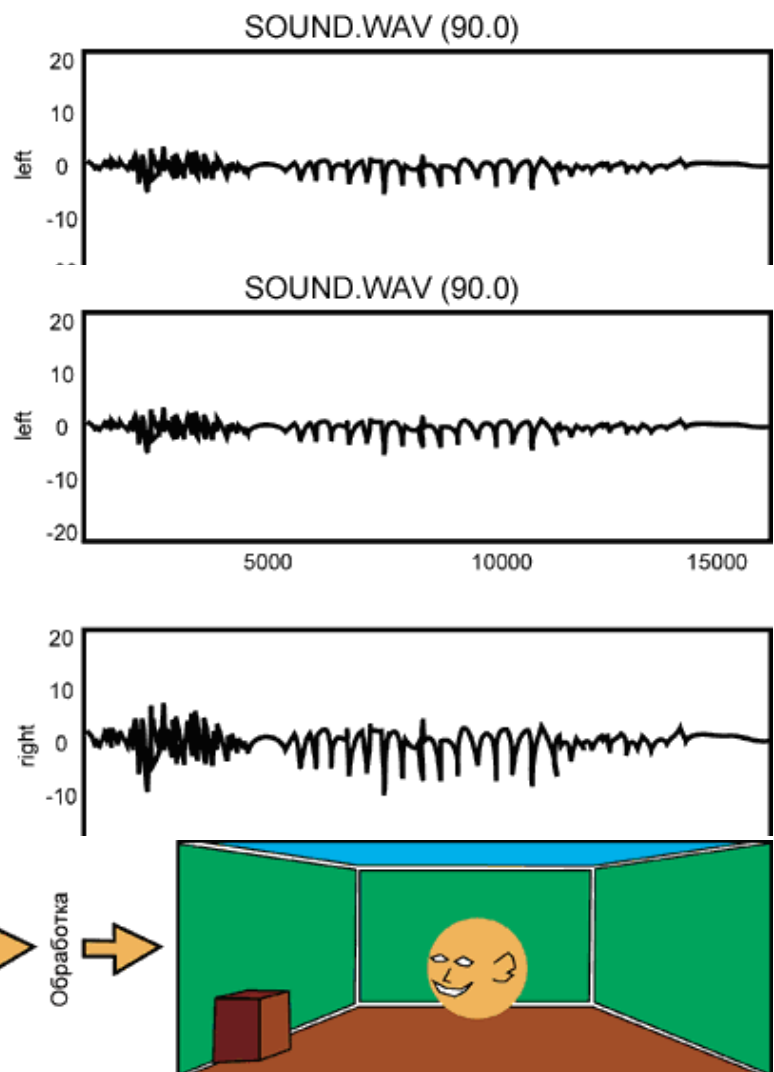
Процесс формирования звуковых сигналов при аурализации проходит следующие последовательные стадии:

- моноуральная запись источника звука
обычно записи делают в заглушенной камере, но допустимо и в

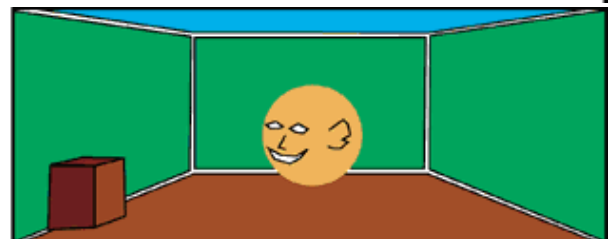
полузаглушенных помещениях, если запись делается там, где отражения не являются определяющими. Расстояние для записи выбирается обычно 1 м на оси. Уровень записи должен соответствовать среднему уровню данного источника звука при естественном звучании. Допускается использование синтезированных музыкальных записей, а также различных банков семплов. Это дает возможность прослушать, как в данном помещении, существующем или желаемом, будут звучать различные виды музыки.

- создание компьютерных моделей звукового поля в помещении как известно из архитектурной акустики, при прослушивании любого звукового источника в помещении к слушателю поступает прямой звук и его многочисленные отражения от стен, потолка, пола и др. (рис.5). Для расчета структуры этих отражений в помещениях различной формы и размеров, с различными поглощающими или отражающими материалами и конструкциями, мебелью, элементами декораций и др., используются пакеты программ, построенные как на приближенных методах геометрической акустики (лучевой метод или метод мнимых источников), так и на более точных методах волновой акустики (МКЭ, МГЭ и др.).

Рис.5. Структура отражений в помещении



Обработка



Если записать микрофоном звуковой сигнал, например, короткий импульс, воспроизведенный через акустическую систему установленную в помещении, то сигнал в любой точке помещения будет иметь вид, показанный на рис.6, т.е. наряду с прямым сигналом в данную точку приходят ранние дискретные отражения, затем число их увеличивается, и процесс приобретает сплошной характер. Важнейшей характеристикой этого реверберационного процесса является время реверберации T_s , т.е. время, в течение которого сигнал затухает на 60 дБ.

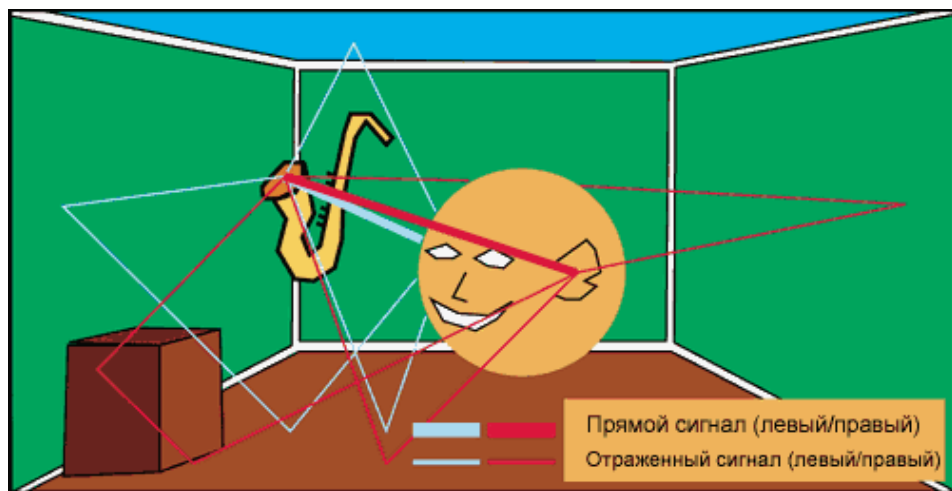
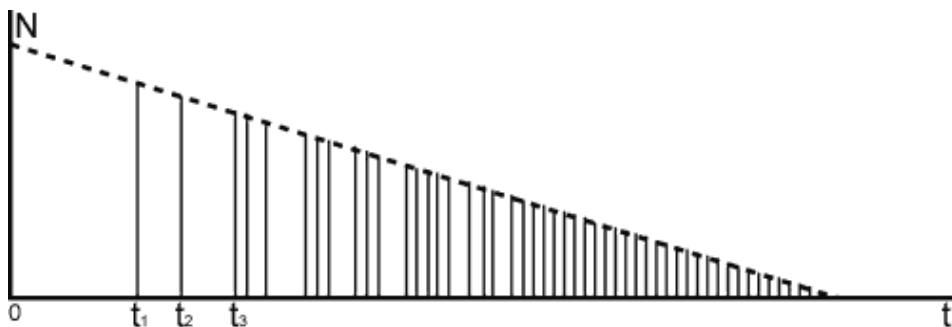


Рис.6. Структура реверберационного процесса в помещении

Время реверберации, структура ранних отражений, характер затухания их на последнем этапе и др. вызывают у слушателя субъективное ощущение размеров пространства, полноты звука, ясности, тембра и др. параметров, по которым отличается акустически хороший зал от плохого. (О связях объективных параметров реверберационного процесса и субъективных ощущениях акустики залов получено за последнее время много новых результатов).



Таким образом для определения импульсных характеристик помещения (под импульсной характеристикой понимается реакция системы на воздействие в виде короткого импульса) можно произвести измерения микрофоном в разных точках пространства, а можно, задав геометрические размеры помещения, поглощающие свойства его стен, потолков и др., рассчитать его для любой точки помещения, что и делается в программах аурализации.

- определение бинауральных импульсных характеристик помещения (BRIR) после того как звуковой сигнал, созданный источником звука, изменен помещением (т.е. к прямому сигналу добавлены его отражения), он обрабатывается двумя слуховыми приемниками, и только после этого он поступает на барабанную перепонку и проходит дальнейшие стадии обработки в периферической слуховой системе и в высших отделах мозга.

Импульсные характеристики, которые получают на входе левого и правого слуховых каналов, называются бинауральными импульсными характеристиками помещения BRIR (binaural room impulse response).

Эти функции BRIR несут в себе всю необходимую информацию: о положении и свойствах источника звука, о свойствах помещения и свойствах приемника звука, то есть обо всех процессах обработки звука, которые происходят в голове, ушных раковинах и др.

Для того чтобы описать эти свойства приемника (т.е. головы и ушных раковин), используются передаточные HRTF (АЧХ и ФЧХ) или импульсные функции слуховой системы - HRIR.

Для определения этих передаточных функций обычно используют библиотеки уже выполненных измерений АЧХ и ФЧХ внутри слухового канала на моделях "искусственной головы" в заглушенной камере при разном расположении источника вокруг головы (Рис.7). Вид передаточных функций существенно меняется (особенно в области частот 5...16 кГц) в зависимости от положения источника в вертикальной и горизонтальной плоскости и по глубине по отношению ко входу в левый и правый слуховые каналы головы. Иначе говоря, ушные раковины, голова и торс действуют как частотно-зависимые дифракционные фильтры. Физические причины этого были рассмотрены в статье о бинауральном слухе (№10/1999).

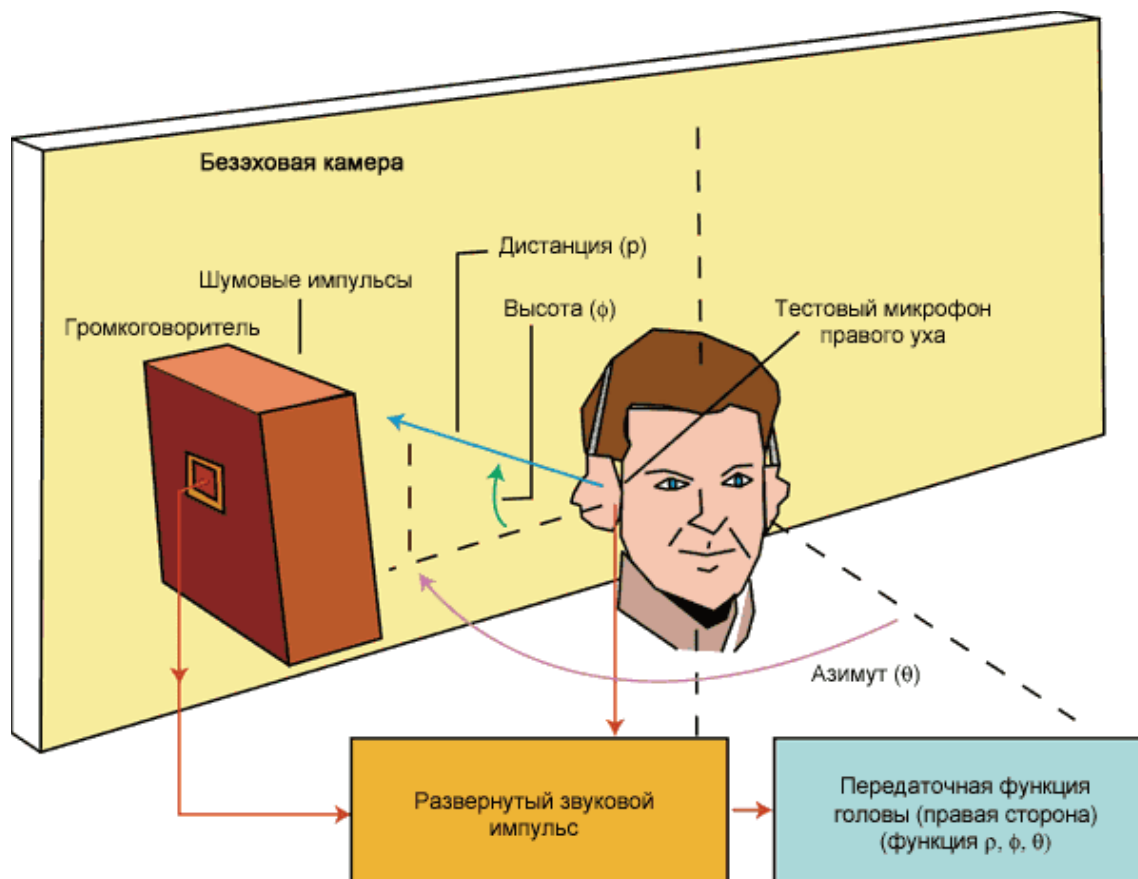


Рис.7. Схема записи передаточных функций слуховой системы при разных положениях источника

В результате для воссоздания пространственного звукового образа необходимо ввести в компьютер следующую информацию:

- выбор источника сигнала: музыка, речь, пение и т.п. Это может быть запись в полузаглушенном или заглушенном помещении или синтезированный сигнал;
- выбор помещения, в которое "помещается" этот источник звука. Могут быть заданы параметры известного помещения или помещения, которое еще предстоит построить;
- положение источника звука в помещении на сцене, на полу, в любой другой точке;
- положение слушателя в каком-либо месте помещения в партере, на балконе и т.д.

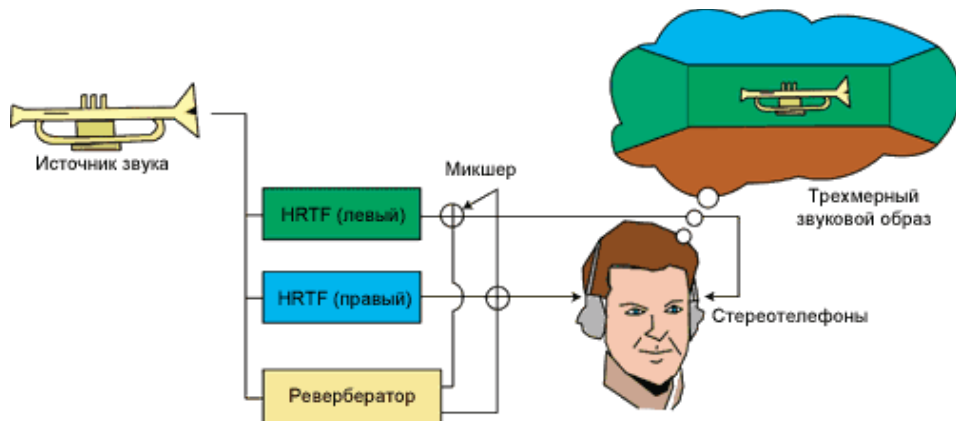
Для реализации "виртуального звукового образа" созданы пакеты компьютерных программ. Наиболее известны программы фирм Odeon и CATT, которые последовательно выполняют следующие операции (Рис.8):

- вводят свойства источника звука из библиотеки записанных или синтезированных звуков в моноварианте;
- производят расчет структуры звукового поля в заданном помещении и вычисляют импульсную характеристику в заданных точках расположения правого и левого ушей слушателя;
- используют из заранее составленной библиотеки значения передаточных функций головы, соответствующих данному положению источника и слушателя;
- производят "свертку", т.е. последовательную обработку фильтрацию сигнала источника с помощью импульсных характеристик помещения и импульсных (передаточных) характеристик головы.

Рис.8. Структура алгоритма

Полученные стереосигналы подают на головные телефоны это дает возможность слушателю почувствовать, что он находится на определенном месте внутри зала, и звук окружает его со всех сторон. При смене положения слушателя или источника производится пересчет передаточных функций.

Как уже было показано в статье, посвященной бинауральному слуху, для нашей слуховой системы существует несколько наиболее важных признаков, по которым она определяет пространственное положение источника. Для локализации в горизонтальной плоскости основное значение имеет разница по времени прибытия сигнала в правое и левое уши, и разница по интенсивности за счет дифракции на голове. Для определения глубины важна разница в уровне звукового давления и разница в спектральном составе, а для локализации в вертикальной плоскости разница в форме АЧХ и ФЧХ за счет дифракции на ушной раковине. Измеренные



значения передаточных функций несут в себе всю необходимую для слуховой системы информацию о локализации источника. Разумеется эти значения передаточных функций сделаны для некоторых усредненных параметров головы и ушных раковин это вносит определенную погрешность, т.к. каждый человек имеет некоторые индивидуальные особенности. Но, во-первых, исследования показали, что погрешности не слишком велики, а, во-вторых, уже созданы компьютерные модели ушной раковины, в которых можно учесть индивидуальные параметры слушателя. Таким образом, компьютерная модель обработки звука, аналогичная работе бинауральных слуховых приемников, должна включать последовательный ряд следующих моделей:

Моноуральная запись => расчет структуры реверберационного процесса (в т.ч. ранние отражения) => модель локализации в вертикальной плоскости (эхо на ушной раковине) => модель локализации глубины (реверберация, громкость) => азимутальная модель локализации (временная и интенсивностная) => выход на левый и правый канал стереотелефонов.

Техника аурализации стремительно развивается - на последней 108-й конвенции AES в Париже было несколько докладов и демонстраций, посвященных программам аурализации с учетом движения головы (система BRS), которая позволяет при поворотах или подъеме головы пересчитать соответствующие параметры звукового поля и дает возможность услышать в наушниках, как соответственно перемещается источник звука в пространстве. Для этого нужен постоянный мониторинг (система обратной связи), которая отслеживает движение головы и пересчитывает соответствующие бинауральные импульсные характеристики. Существуют разные системы обратной связи, с помощью которых это можно делать: от простейших инфракрасных датчиков, с использованием которых уже давно выпускаются стереотелефоны, до сложных и дорогостоящих систем обратного контроля. Когда эта технология окончательно созреет, качество систем пространственного звуковоспроизведения перейдет на новый уровень.

На 108-й конвенции AES были представлены специальные демонстрационные системы фирмы Studer, когда слушатель мог через мониторы сначала прослушать пространственный звук, который при этом перемещался от одной акустической системы к другой (в комнате были установлены передние, задние, боковые и центральные системы). Затем с помощью компьютерной системы аурализации производилось прослушивание записей через головные телефоны с системой обратной связи.

При этом можно было услышать полную пространственную картину внешнего окружающего звукового поля, которое перемещалось при повороте головы это действительно впечатляет!

Конечно, хотелось бы использовать в системах аурализации не только головные телефоны, но и иметь возможность прослушивать обработанные записи через акустические системы. Однако для этого надо решить еще несколько дополнительных проблем: во-первых, вторичное помещение накладывает свое реверберационное поле, что в данном случае является помехой, ведь вся необходимая информация о помещении уже закодирована в сигнале. Во-вторых, сигналы от левого канала попадают не только на левое ухо, но и на правое, т.е. возникают перекрестные связи, которые разрушают звуковой образ.

В настоящее время активно развивается техника бинауральной стереофонии, в которой необходимо решение тех же проблем. За последние годы разработаны различные методы проектирования бифонических процессоров, которые реализуют подавление перекрестных связей в реверберирующих помещениях; и адаптивных процессоров, которые могут подавлять отражения во вторичном помещении. Правда, осталась проблема расширения зоны прослушивания, так как пока удается это сделать для фиксированной позиции слушателя. (О бинауральной стереофонии постараюсь рассказать в дальнейшем). Когда будет достигнут прогресс и этих направлениях, можно ожидать, что появится возможность прослушивания через акустические системы. Следует отметить, что достаточно будет двух акустических систем для левого и правого каналов для воссоздания пространственного звучания это-то и заманчиво!

Новую технологию компьютерного создания пространственных звуковых полей "аурализацию" несомненно, ожидает много применений:

- в архитектурной акустике-для оценки акустических свойств различных существующих залов и моделирования еще не построенных помещений, для оценки влияния различных элементов звукопоглощающих конструкции на качество звучания, в том числе и студий звукозаписи, для проектирования систем звукоусиления и др.;
- в технике звукозаписи открывается много новых возможностей для звукорежиссеров в создании пространственных эффектов, совершенствовании систем пространственной звукопередачи;
- в системах мультимедиа для создания "виртуальных реальностей" как видео-, так и звуковых и не только в компьютерных играх;
- для обучения музыкантов, певцов, артистов для получения возможности предварительного прослушивания различных видов исполнения в залах с разной акустикой;
- для тренировки слепых в распознавании и локализации источников звука;
- в постановке научных экспериментов, в частности, в психоакустике;
- в системах обнаружения и распознавания различных источников сигнала в пространстве (в авиации и др.)

По мере развития этой технологии, несомненно, появятся новые применения. Хотелось бы пожелать, чтобы она нашла себе широкое применение и в отечественной звукотехнике.

Часть 11. Громкость, ч.1

Как уже было отмечено в предыдущих статьях по психоакустике, звуковой сигнал (музыка, речь, шум и др.), поступающий на вход слуховых каналов, вызывает у слушателя определенные субъективные ощущения, основными из которых являются высота звука, громкость, тембр, пространственность и др. Каждое из этих ощущений сложным и неоднозначным образом связано с объективными параметрами звукового сигнала: интенсивностью, длительностью, спектральным составом, локализацией в пространстве и др. Установление этих связей и определение количественных соотношений между ними и есть одна из основных задач психоакустики.

Человеческий слух обладает удивительной способностью реагировать на слуховые сигналы как очень малой интенсивности (звуковое давление 2×10^{-5} Па - уровень 0 дБ), так и очень большой интенсивности (звуковое давление 20 Па - уровень 120 дБ), это соответствует динамическому диапазону 120 дБ.

Громкостью называется субъективное ощущение, позволяющее слуховой системе располагать звуки по определенной шкале от звуков низкой интенсивности ("тихие" звуки) к звукам большой интенсивности ("громкие" звуки).

Громкость связана прежде всего с таким физическим параметром звукового сигнала как его интенсивность (т.е. звуковая энергия). Интенсивность I и звуковое давление p связаны простым (для плоской волны) соотношением $I = p^2 / \rho c$, где ρ - плотность воздуха, c - скорость звука.

Общеизвестно, что чем больший уровень звукового давления (дБ) создает акустическая аппаратура, тем она громче звучит. Однако все далеко не так просто - можно создать звуковые сигналы очень большой интенсивности, и при этом никакого ощущения громкости не вызвать. И это при том, что слуховая система может быть даже повреждена - например, в случае, если эти сигналы будут слишком короткими (менее 35 мс) или слишком низкочастотными (ниже 100 Гц).

Происходит это потому, что громкость зависит не только от интенсивности звука, но и от его частоты, спектрального состава, длительности, локализации в пространстве и др.

Громкость звука - это субъективная величина, она характеризует ощущение слушателя, поэтому громкость не может быть измерена прямыми методами. Возможно, в ближайшем будущем это можно будет сделать на компьютерных моделях слуховой системы, которые сейчас усиленно развиваются.

В настоящее время оценки ощущения громкости при изменении различных параметров звукового сигнала получаются методом субъективных экспертиз: либо сравнением с эталонным звуком, либо абсолютной оценкой. Процессы эти очень трудоемки, требуют проведения большого количества экспериментов, накопления статистических данных. Исследования процессов ощущения громкости все время продолжаются в ведущих научных институтах, как отечественных так и зарубежных. Постоянно появляются публикации об уточнении известных соотношений и о новых результатах. Наиболее известные ученые, труды которых используются в этом направлении - Бекеши, Стивенс, Цвиккер, Гельфанд, Мур.

Понимание механизмов ощущения громкости и ее зависимости от основных объективных параметров звукового сигнала имеет чрезвычайно важное значение для практики работы звукорежиссеров - так, например, если запись музыкального произведения и его прослушивание происходит на разных уровнях интенсивности, то ощущение баланса громкости и, следовательно, тембра звучания будет совершенно разным у звукорежиссера и у слушателя, что следует учитывать при записи и при воспроизведении.

Шкалирование звуков по громкости и установление влияния на него основных параметров сигнала было выполнено, прежде всего, для тональных сигналов различной интенсивности, частоты и длительности, что послужило основой для оценки громкости сложных музыкальных, речевых и шумовых сигналов.

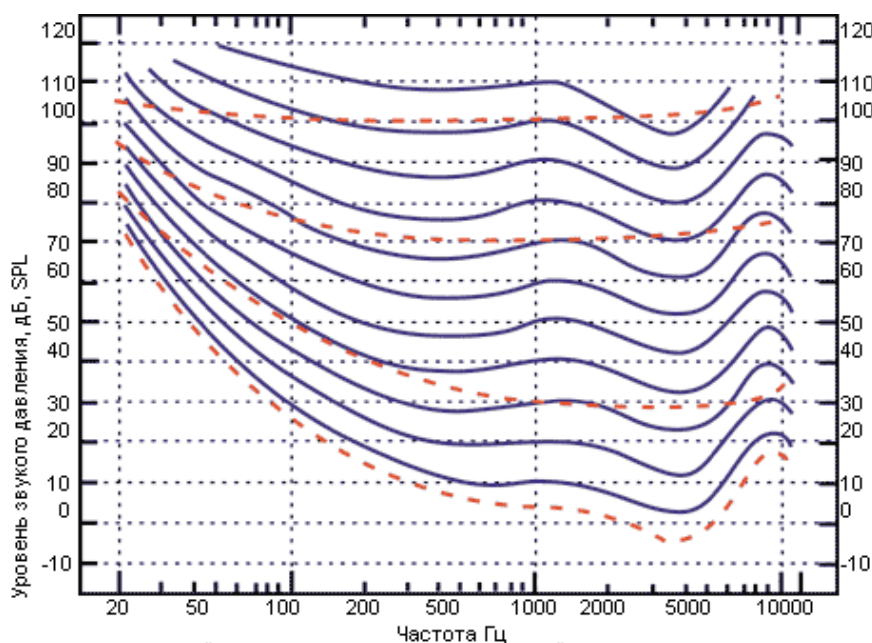


Рис 1 Кривые равной громкости для различных значений уровня сигнала

Поскольку техника оценки абсолютной громкости и ее связей с интенсивностью, частотой и длительностью звуковых сигналов достаточно сложна, то широкое распространение получили методы относительной оценки уровней громкости. Уровни громкости определяются с помощью экспериментов.

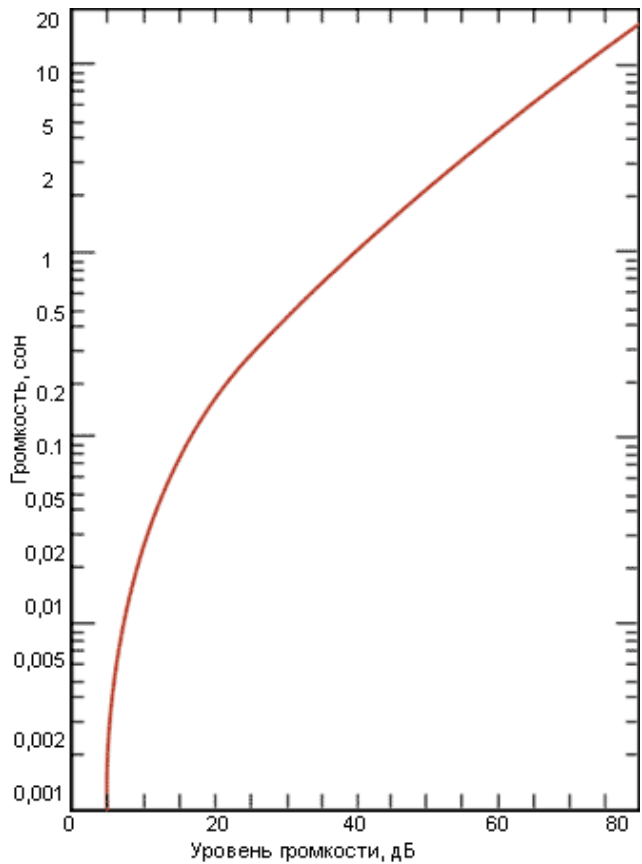


Рис 3 Связь между громкостью и уровнем громкости

Выставляется уровень звукового давления эталонного звука на частоте 1000 Гц (например, 40 дБ), затем испытуемому предлагается прослушать сигнал на другой частоте (например, 100 Гц), и отрегулировать его уровень таким образом, чтобы он казался равногромким эталонному. Сигналы могут предъявляться через телефоны или через громкоговорители. Если проделать это для разных частот, и отложить полученные значения уровня звукового давления, которые требуются для сигналов разной частоты, чтобы они были равногромкими с эталонным сигналом - получится одна из кривых на рис. 1.

Например, чтобы звук с частотой 100 Гц казался таким же громким, как звук с частотой 1000 Гц с уровнем 40 дБ, его уровень должен быть выше, около 50 дБ. Если будет подан звук с частотой 50 Гц, то, чтобы сделать его равногромким с эталонным, нужно поднять его уровень до 65 дБ и т.п. Если теперь увеличить уровень эталонного звука до 60 дБ и повторить все эксперименты, то получится кривая равной громкости, соответствующая уровню 60 дБ...

Семейство таких кривых для различных уровней 0, 10, 20...110дБ показано на рис. 1. Эти кривые называются кривыми равной громкости. Они были получены учеными Флетчером и Мэнсоном в результате обработки данных большого числа экспериментов, проведенных ими среди нескольких сотен посетителей Всемирной выставки 1931 года в Нью-Йорке.

В настоящее время в международном стандарте ISO 226 (1987 г.) приняты уточненные данные измерений, полученные в 1956году. Именно данные из стандарта ISO и представлены на рис.1, при этом измерения выполнялись в условиях свободного поля, то есть в заглушенной камере, источник звука располагался фронтально и звук подавался через громкоговорители. Сейчас накоплены новые результаты, и предполагается в ближайшем будущем уточнение этих данных. Каждая из представленных кривых называется изофоной и характеризует уровень громкости звуков разной частоты.

Под уровнем громкости данного звука понимается уровень звукового давления равногромкого с данным эталонного звука на частоте 1000 Гц. Уровень громкости измеряется в специальных единицах - фонах.

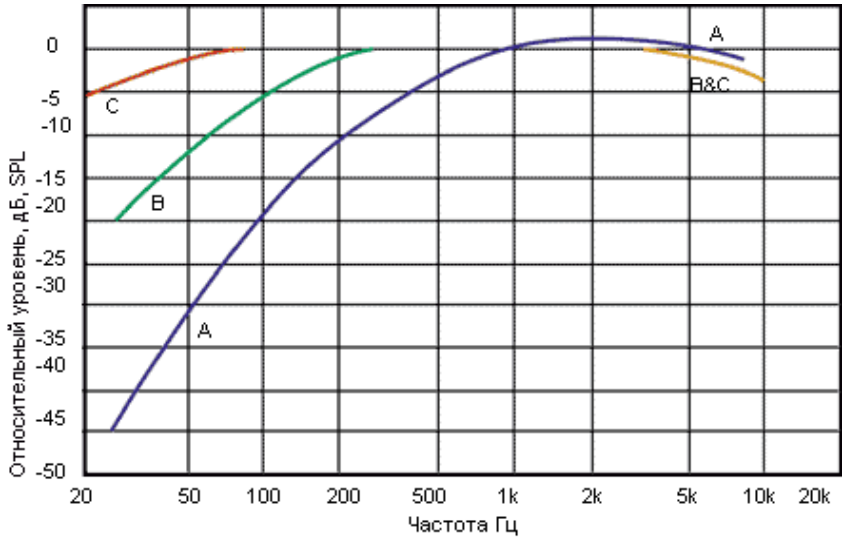


Рис 2 Взвешенные кривые A,B,C для измерения уровней шума

Цифры, которые стоят над кривыми на рисунке 1, соответствуют числу фон, которые равны числу децибел звука с частотой 1000 Гц. Зная частоту данного сигнала и его уровень звукового давления, можно, пользуясь представленными кривыми, определить его уровень громкости в фонах. Например, если тональный звук с частотой 100 Гц имеет уровень звукового давления 60 дБ, то, проведя прямые, соответствующие этим значениям на рис. 1, находим на их пересечении изофону, соответствующую уровню 50 фон, - значит, этот звук имеет уровень громкости 50 фон.

Если проанализировать эти кривые, то видно, что при малых уровнях звукового давления оценка уровня громкости очень сильно зависит от частоты - слух менее чувствителен к низким и высоким частотам, и требуется создать гораздо большие уровни звукового давления, чтобы звук стал звучать равногромко с эталонным звуком 1000 Гц. При больших уровнях изофоны выравниваются, подъем на низких частотах становится менее крутым - происходит более быстрое нарастание громкости звуков низкой частоты, чем средних и высоких. Таким образом, при больших уровнях низкие, средние и высокие звуки оцениваются по уровню громкости более равномерно.

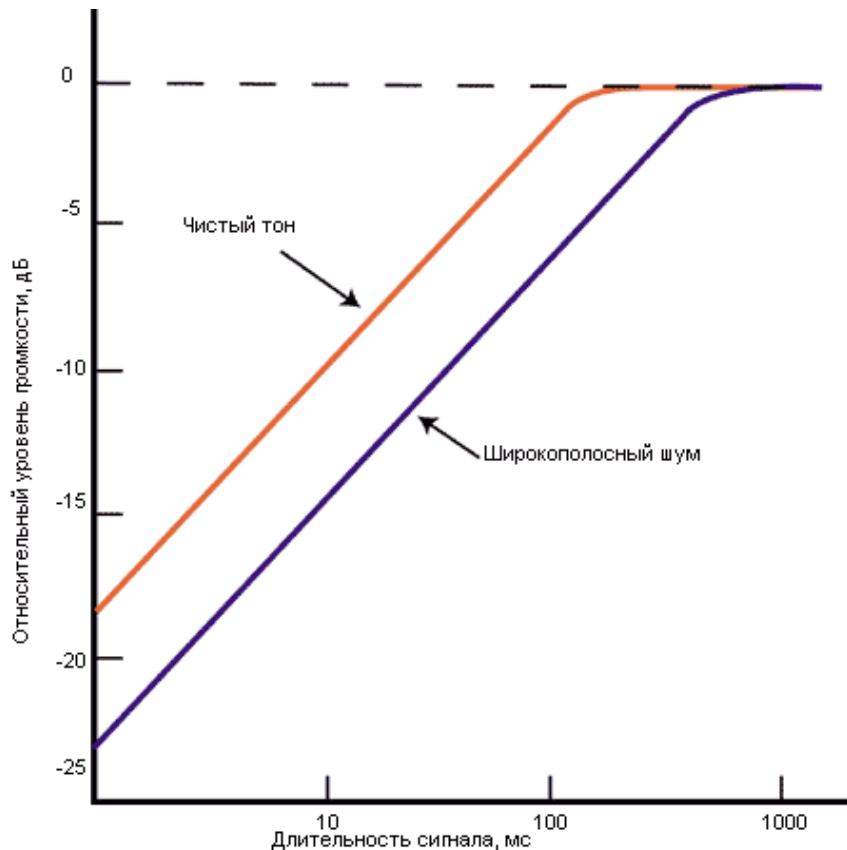


Рис 4 Возрастание уровня громкости при увеличении длительности сигнала

Это свойство слуха имеет огромное значение для техники звукозаписи, т.к. относительная громкость звуков разной частоты изменяется при изменении общего уровня записи, но только в том случае, если запись воспроизводится на том же уровне, что и оригинальный источник. Тогда в ней сохраняется естественный баланс по громкости. Если же запись воспроизводится на более низких уровнях, низкие и высокие частоты как бы пропадают, баланс нарушается. Это одна из причин того, почему пение и речь кажутся бубнящими при воспроизведении на высоких уровнях через громкоговорители: слушатель воспринимает в них низкие частоты значительно более громкими, чем при прослушивании естественного источника на более "тихих" уровнях.

Это свойство слуховой системы - по-разному оценивать уровень громкости сигнала в зависимости от его частоты и уровня звукового давления - учитывается в современных приборах для измерения уровней шума и других сложных звуков. В них применяются взвешенные корректирующие кривые, аналогичные кривым "слуховых" фильтров, которые ослабляют низкие частоты в зависимости от уровня сигнала так, как это делает слуховая система.

Обычно используются три вида взвешенных кривых (рис.2):

- кривая А со спадом -30 дБ на уровне 50 Гц по отношению к уровню на 1000 Гц;
- кривая В со спадом -12 дБ;
- кривая С со спадом -2 дБ.

Если эти кривые перевернуть "сверху вниз" (зеркально по горизонтальной оси), и нанести на изофоны (пунктирные линии на рисунке 1), то видно, что кривая А примерно соответствует изофоне 30 фон. Таким образом, проводя измерения с использованием этой кривой (значения уровня сигнала выдаются в дБА) мы как бы оцениваем уровень громкости этого сигнала так, как это делает слуховая система на слабых уровнях (30 дБ на 1000 Гц), кривая В (значения в дБВ) соответствует изофоне 70 дБ, кривая С - изофоне 100 дБ (дБС).

Разумеется, использование таких приборов позволяет оценить уровень громкости только очень приблизительно, т.к. при оценке громкости сложных сигналов слуховая система использует более сложные механизмы, о которых поговорим дальше.

Оценка уровня громкости не эквивалентна оценке изменения абсолютной громкости. Например, если имеется два сигнала с уровнями громкости 40 и 80 фон, то это не значит, что один громче другого в два раза. Связь уровня громкости с абсолютной оценкой громкости носит достаточно сложный характер.

Наряду с созданием методов сравнительной оценки уровня громкости, постоянно продолжаются попытки построения шкалы для оценки абсолютного ощущения громкости в зависимости от интенсивности, частоты, длительности и других объективных параметров сигнала. Задача эта намного сложнее, чем предыдущая - достаточно сравнить методы оценки качества звучания методом парного сравнения (для этого в качестве экспертов можно привлекать широкий круг слушателей) и оценку качества звучания по абсолютной шкале - это доступно только экспертам с большим музыкальным опытом, например, музыкантам и звукорежиссерам.

Обычно для решения такого рода задач (в наиболее известных работах Стивенса) использовались два метода.

Первый: испытуемым предъявлялись звуки разной интенсивности, и их просили присвоить численно оценить (в баллах) каждый звук в зависимости от воспринимаемой громкости.

Второй метод: подавался эталонный звук, и испытуемых просили оценить громкость измеряемого звука относительно заданного, например, в два раза, в три раза и т.д. Обработка большого количества статистических данных позволила построить графики зависимости ощущаемой громкости от уровня звукового давления. Для количественной оценки абсолютной громкости была принята специальная единица сон. Громкость в 1 сон - это громкость синусоидального звука с частотой 1000 Гц и уровнем 40 дБ.

Количественно зависимость воспринимаемой громкости звука (в сонах) и его звукового давления (в Па) может быть представлена в следующем виде:

$S = C \times p^{0,6}$, где C - постоянная, зависящая от частоты сигнала.

Из этого соотношения следует, что зависимость является нелинейной, что подтверждает общий закон психофизики о том, что зависимость между изменением объективных параметров сигнала и возникающими при этом субъективными ощущениями носит нелинейный логарифмический характер. Из этого соотношения получается также, что при увеличении уровня звукового давления на 10 дБ громкость возрастает в два раза. Например, если на частоте 1000 Гц сигнал с уровнем 40 дБ создает ощущение громкости в 1 сон, то сигнал с уровнем 50 дБ соответствует громкости в 2 сона, 60 дБ - 3 сона и т. д.

Правда, в некоторых исследованиях установлено, что удвоение громкости вызывается увеличением уровня сигнала только на 6 дБ - особенно для низких частот.

Аналогично связи между значением звукового давления в Па и его уровнем в дБ, между абсолютным значением громкости в сонах S и значением уровня громкости в фонах L существует связь (в стандартах ISO): $S = 2^{L-40/10}$.

Графическая зависимость громкости в сонах от уровня громкости в фонах для частоты 1000 Гц показана на

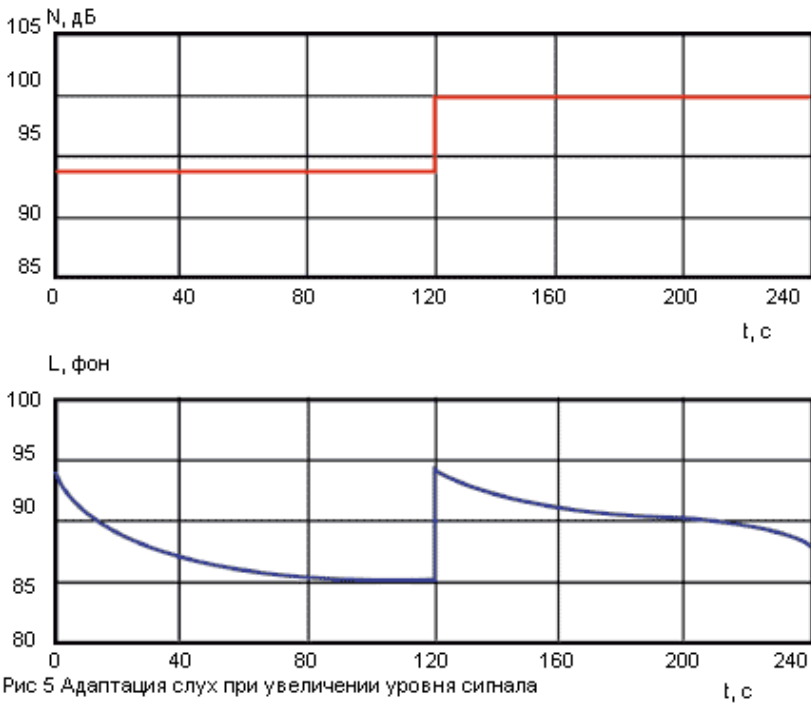


Рис 5 Адаптация слуха при увеличении уровня сигнала

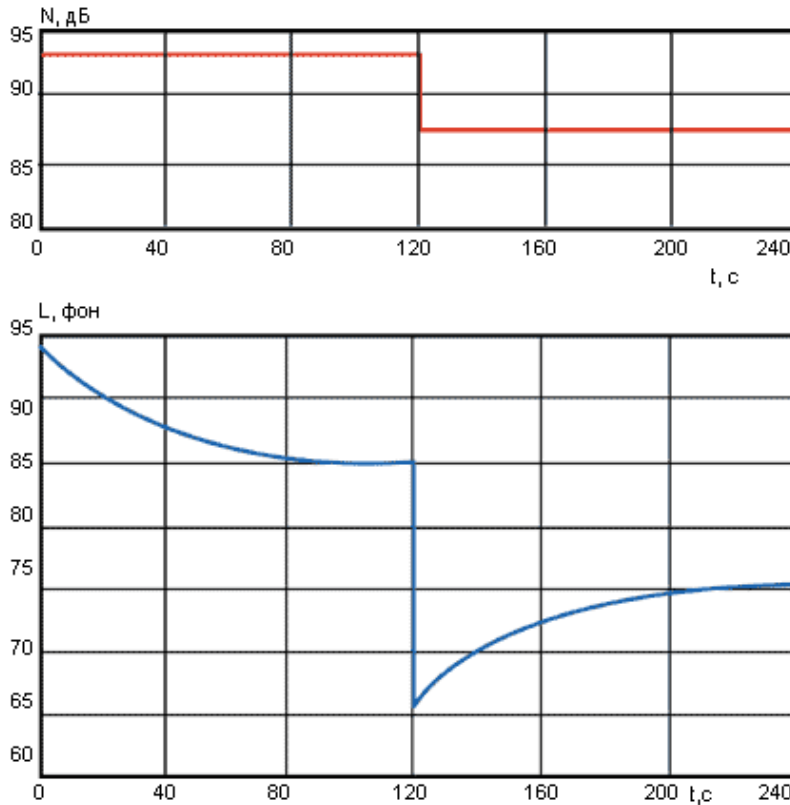


Рис 6 Адаптация слуха при уменьшении уровня сигнала

рисунке 3. Эта зависимость построена для измерений, выполненных в свободном поле при прослушивании через громкоговорители. Полученные количественные соотношения очень важны для определения громкости сложных звуков, которые будут рассмотрены дальше.

Результаты шкалирования абсолютных значений громкости сильно зависят от ряда факторов: индивидуальных слуховых различий, порядка предъявления стимулов, тренированности и концентрации внимания экспертов. Поэтому для получения значимых результатов требуется большое количество экспериментов, и работы в этом направлении, как уже сказано, постоянно продолжаются. Следует отметить, что при количественной оценке громкости реальных сложных сигналов процесс зависит не только от временной и спектральной структур сигнала, но и от его смыслового содержания и окружающей пространственной обстановки.

Прежде, чем переходить к анализу громкости сложных звуков, остановимся еще на двух существенных моментах.

Ощущение громкости зависит от длительности сигнала: если на слуховой канал поступают два сигнала одинаковой интенсивности, то более короткий сигнал, воспринимается как менее громкий. Это полезно учитывать при обработке музыкальных и речевых сигналов. При увеличении длительности сигнала ощущение громкости постепенно возрастает, пока его длительность не достигает величины 100...200 мс, при этом возрастание уровня громкости происходит почти линейно с увеличением длительности сигнала (рис. 4).

Слуховая система обладает свойством адаптации, т.е. под воздействием длительных, громких, постоянных по величине звуков ощущаемая громкость звука постепенно уменьшается - слух адаптируется. Результаты изменения уровня звукового давления и ощущаемого уровня громкости (полученные знаменитым ученым Бекешем) показаны на рисунке 5.

При воздействии звука с уровнем 94 дБ в течение двух минут уровень громкости постепенно уменьшается на величину 9 фон, при этом к концу периода времени падение замедляется. Если при этом уровень сигнала резко увеличить, например с 94 до 100 дБ, то уровень громкости увеличивается, однако в меньшей степени, чем это должно было бы соответствовать значению уровня сигнала в 100 дБ. Затем уровень громкости начинает опять снижаться, и даже с большей скоростью, т.е. степень адаптации тем больше, чем громче звуковой сигнал. При этом происходит снижение чувствительности слуха и повышение слуховых порогов (о чем было рассказано в предыдущей статье).

Изменение уровня громкости проявляется и при внезапном уменьшении уровня воздействующего сигнала. Как показано на рисунке 6, при воздействии сигнала с уровнем 94 дБ происходит постепенная адаптация (как и предыдущем случае, на 9 дБ), затем, при скачкообразном уменьшении уровня сигнала на 6 дБ, уровень ощущаемой громкости резко падает на 19 фон, а затем постепенно увеличивается, т.е. происходит адаптация к тихим звукам, и постепенно чувствительность восстанавливается.

Таким образом, слуховая система пытается защититься от громких звуков - при их длительном воздействии происходит постепенное снижение ощущения громкости, звуки кажутся более тихими. Степень адаптации зависит от громкости воздействующего сигнала - чем он громче, тем больше снижение ощущаемого уровня громкости. Однако возможности слуховой системы ограничены, и процесс имеет тенденцию к насыщению: например, при переходе от уровня 94 дБ к уровню 108 дБ разница в снижении уровня ощущаемой громкости происходит всего на 3 фона.

В основе процесса адаптации лежат механизмы, происходящие в среднем и внутреннем ухе. В статье, посвященной анализу высоты музыкального звука, был показан механизм работы среднего и внутреннего уха, при этом отмечено, что на больших уровнях сигнала срабатывает так называемый "акустический рефлекс". При этом стремечко отводится от овального окна и предохраняет внутреннее ухо от передачи слишком громких звуков. Рефлекс начинает срабатывать для звуков с уровнем 85 дБ и выше, и обеспечивает защиту до 20 дБ. Кроме того, процесс колебаний базилярной мембраны является сугубо нелинейным - при слишком больших смещениях мембраны происходит компрессия сигнала за счет действия наружных волосковых клеток.

Однако защитная способность слуха, как уже сказано, ограничена; кроме того, она обладает определенной инерцией - акустический рефлекс начинает срабатывать только через 30...40 мс после начала звука, и полная защита еще не достигается и при 150 мс, поэтому, наряду с опасностью для слуховой системы воздействия длительных громких звуков, еще более опасным для нее является воздействие коротких громких импульсов.

Таким образом, ощущение громкости сложным нелинейным образом зависит от интенсивности воздействующего сигнала, его частоты и длительности. Однако еще большую проблему представляет определение зависимости ощущаемой громкости от спектрального состава сложных музыкальных и речевых сигналов, что будет рассмотрено во второй части данной статьи.

Часть 12 Громкость сложных звуков, часть 2

Как было показано в первой части статьи, посвященной субъективному восприятию громкости звука ("Звукорежиссер", 8/2000), ощущение громкости, как меры распределения звуков от тихих до громких по определенной шкале, зависит от таких объективных параметров, как интенсивность звука (звуковое давление), частота, длительность, спектр, маскирующее действие других звуков и др. Зависимость уровня громкости (выраженного в фонах) от частоты была представлена в виде кривых равной громкости. Там же была рассмотрена для простых тональных звуков зависимость громкости (в сонах) от уровня громкости, от длительности сигнала и пр.

Перейдем теперь к анализу восприятия громкости для сложных звуков, т.е. рассмотрим зависимость ощущения громкости от спектрального состава различных сигналов (речевых, музыкальных, шумовых и др.), что особенно важно учитывать на практике при записи, монтаже и других видах работ со звуковым материалом.

Известно из практики, что широкополосные сигналы кажутся громче, чем узкополосные сигналы с таким же уровнем звукового давления. Пример зависимости уровня громкости от ширины полосы шумового сигнала показан на рисунке 1 для уровня звукового давления 60 дБ и центральной частоты 1 кГц. Воспринимаемый уровень громкости при расширении полосы шума до определенного значения (в данном случае 150 Гц) практически остается неизменным, а когда ширина полосы становится шире 150 Гц, уровень громкости резко возрастает. Граница, где происходит изменение ощущения уровня громкости, называется критической полосой слуха. Различие механизмов обработки сигнала внутри и вне критических полос имеет принципиальное значение для определения громкости сложных звуков (так же, как их высоты, тембра и др.).

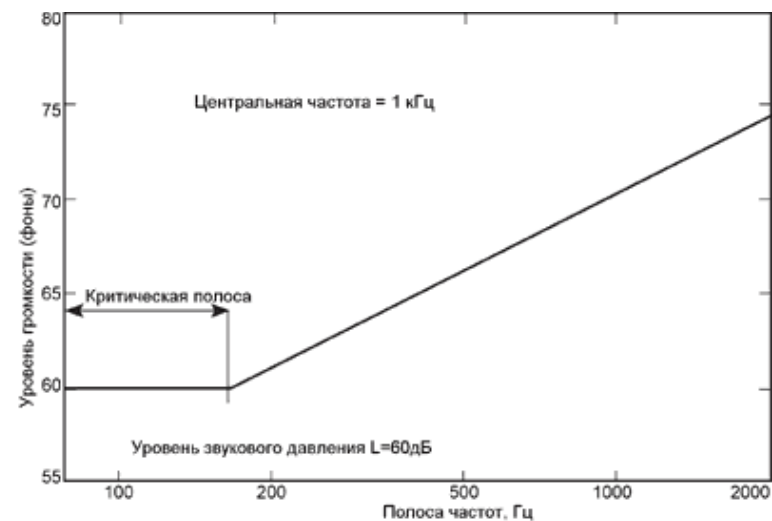


Рис. 1. Зависимость уровня громкости от ширины полосы сигнала

Механизм ощущения громкости продолжает оставаться предметом многочисленных исследований психоакустиков, однако расшифровка этого процесса по-прежнему представляет значительные трудности.

Одна из самых последних компьютерных моделей слухового анализа громкости сигналов, выполненная учеными Кембриджского Университета (Б. Мур, Б. Гласберг и др.), включает в себя следующие последовательные этапы обработки звукового сигнала в процессе формирования ощущения громкости:

- фильтрация сигнала внешним ухом (ушной раковиной и слуховым каналом);
- фильтрация сигнала средним ухом;
- фильтрация с помощью линейки полосовых фильтров на базилярной мембране;
- преобразование возбуждения на базилярной мембране в кривые распределения удельной громкости;
- интегрирование площади под кривыми удельной громкости.

Поговорим об этих этапах подробнее.

Рис. 2. АЧХ сигнала на барабанной перепонке после фильтрации внешним ухом

Как уже было сказано в предыдущих статьях о бинауральном слухе, внешнее ухо производит обработку звукового сигнала в зависимости от направления прихода звука, увеличивая уровень сигнала в области 3 кГц за счет дифракции на ушной раковине и резонансов в наружном слуховом канале. Равномерная форма АЧХ сигнала после фильтрации внешним ухом приобретает при осевом падении звукового сигнала вид, показанный на рис.2.

Как уже было показано в статье по определению высоты тона, во внутреннем ухе происходит спектральный анализ поступившего слухового сигнала, при этом каждой частоте соответствует свое место максимального смещения базилярной мембраны, что аналогично механизму обработки сигнала линейкой полосовых ("слуховых") фильтров. Ширина критических полос примерно соответствует ширине полосы пропускания слуховых фильтров и меняется в зависимости от частоты в соответствии с кривой на рисунке 3 (для сравнения приведено изменение ширины полосы, соответствующей третьоктавной полосе и целому музыкальному тону).

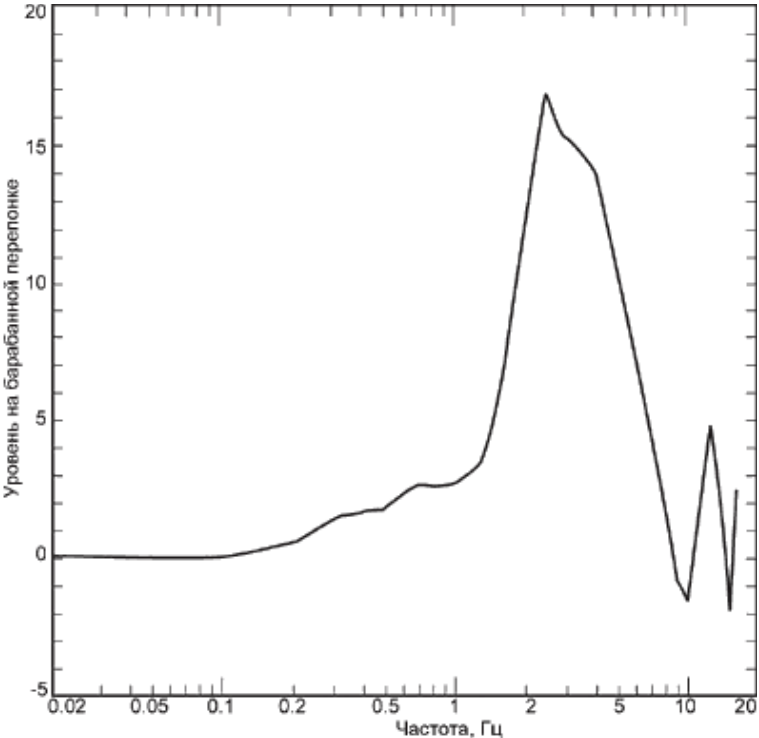
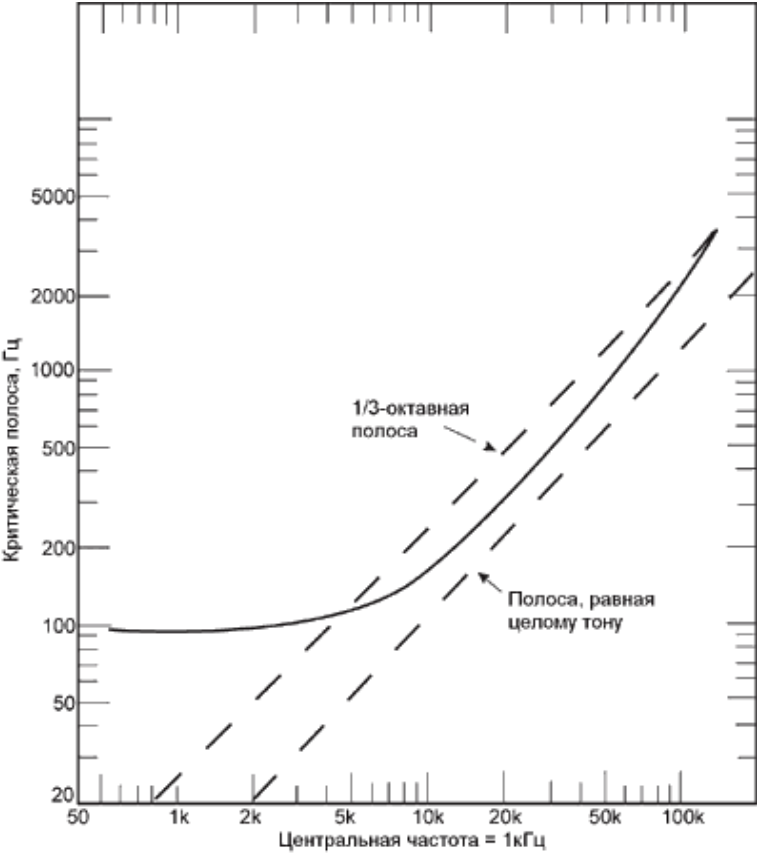


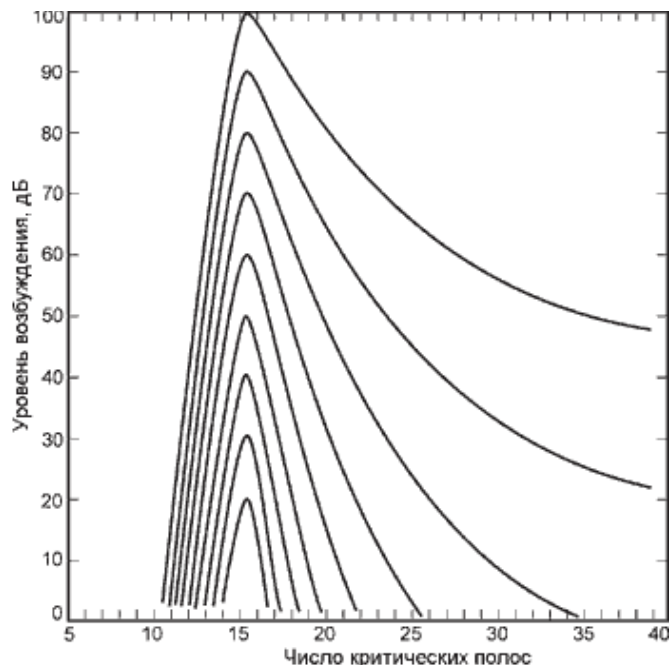
Рис. 3. Изменение ширины критических полос как функции частоты

Если совместить критические полосы в один ряд, то в слышимом диапазоне их оказывается 24, каждой из них соответствует расстояние на базилярной мембране, равное 1,3 мм. Переход от одной критической полосы к другой соответствует изменению высоты в 100 мел или в 1 барк. Форма передаточной функции каждого из этих слуховых фильтров, и ее изменение с увеличением амплитуды сигнала показаны на рисунке 4 (по горизонтальной оси отложено число критических полос). Как видно из рисунка, возбуждение мембраны становится все более несимметричным, и площадь под кривой расширяется.



При колебаниях базилярной мембраны, в волосковых клетках органа Корти (находящегося на мембране) генерируется электрический потенциал (подробнее см. "Основы психоакустики", ч.1), и возбуждаются потоки импульсов в нервных клетках. При увеличении интенсивности сигнала скорость импульсов увеличивается в единичном нерве, соответствующем данному месту на мембране, и доходит до насыщения (порог - 1000 импульсов в секунду), затем начинает возникать возбуждение в соседних нервных волокнах в соответствии с увеличением площади под кривой возбуждения (Рис. 4).

Рис. 4. Изменение формы АЧХ слухового фильтра с изменением интенсивности сигнала



Ощущение громкости кодируется увеличением числа разрядов в единичном нерве и увеличением количества нервных волокон, в которых возникает возбуждение, в соответствии с изменением площади под кривой возбуждения.

На основании проведенных экспериментов была высказана гипотеза, что слуховая система производит интегрирование площади под кривой возбуждения на базилярной мембране с учетом распределения нервной активности.

В упомянутой ранее компьютерной модели слуховой оценки громкости были предложены количественные соотношения между энергией подводимого сигнала, распределением ее по критическим полосам слуха, и возникающим при этом субъективным ощущением удельной громкости в сонах. Под "удельной" понимается оцениваемая громкость внутри критической полосы.

Таким образом, внутри каждой критической частоты происходит интеграция энергии независимо от вида звукового сигнала. Фрагмент спектра шума (или тональные сигналы), если они находятся внутри критической полосы и имеют

одинаковый уровень интенсивности (звукового давления), создают одинаковый уровень громкости.

Поэтому, когда звуковой сигнал имеет сложный спектральный состав или одновременно звучат несколько сигналов, определение их суммарной громкости происходит тремя различными способами, в зависимости от соотношения их частот или обертонов:

- если сигналы близки по частоте, т.е. находятся внутри критической полосы, то для определения создаваемой им суммарной громкости необходимо сложить их интенсивности $I = I_1 + I_2 + I_3 \dots$ и по суммарному значению уровня звукового давления, соответствующего этой суммарной интенсивности, определить из кривых равной громкости уровень громкости (в фонах), а затем пересчитать в значение громкости в сонах.

Рис. 5. Кривые равной громкости

Например, если на скрипке исполняется определенная нота с уровнем звукового давления $L_1 = 60$ дБ с частотой 880 Гц (Ля второй октавы), это, как следует из кривых равной громкости (Рис. 5), соответствует уровню громкости $L_s = 60$ фон. Для определения громкости этого звука можно воспользоваться стандартным соотношением, рекомендованным международными стандартами ISO: $S = 2(L_s - 40)/10(1)$, откуда громкость S равна четырем сонам.

Если теперь будут вместе играть десять скрипок, то создаваемая ими громкость определяется следующим образом: интенсивность звука одной скрипки I_1 , интенсивность звука десяти скрипок $I_{сум} = 10I_1$ (интенсивности складываются).

При этом суммарный уровень интенсивности равен:

$$10 \lg I_{сум}/I_0 = 10 \lg 10I_1/I_0 = 10 \lg I_1/I_0 + 10 \lg 10 = 10 \lg I_1/I_0 + 10 \text{ дБ(2)}.$$

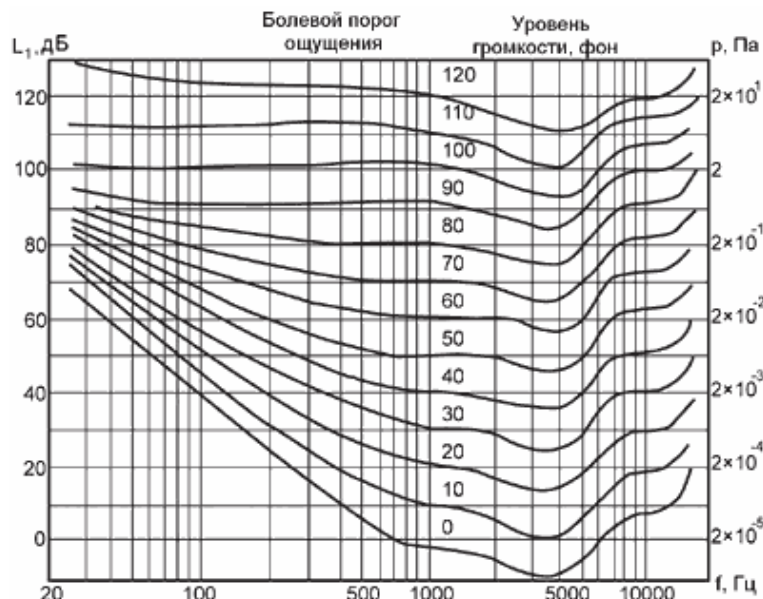
Если теперь учесть, что интенсивность звука пропорциональна квадрату звукового давления, то получим: $10 \lg I_{сум}/I_0 = 10 \lg p_2^2/p_0^2 = 20 \lg p_{сум}/p_0$.

Из соотношения (2) получается: $20 \lg p_{сум}/p_0 = 20 \lg p_1/p_0 + 10 \text{ дБ}$, т.е. суммарный уровень звукового давления увеличится на 10 дБ: $L_p = L_1 + 10 \text{ дБ}$.

Поскольку начальный уровень звукового давления был 60 дБ, то суммарный уровень звукового давления будет 70 дБ, что соответствует уровню громкости 70 фон (рисунок 5), отсюда по формуле (1) можно рассчитать громкость, она равна 8 сон.

Следовательно, когда вместо одной скрипки (или любого другого источника сигнала) будут играть десять скрипок, громкость вырастет только в два раза (от 4 сон до 8 сон), что очень важно учитывать в технологии звукозаписи.

Это правило можно сформулировать иначе: при увеличении общего уровня звукового давления на 10 дБ



воспринимаемая громкость удваивается.

Аналогично рассчитывается общая громкость, если имеется два узкополосных шума с близкими частотами (например, внутри критической полосы около 1000 Гц ее ширина равна 150 Гц). Если уровень каждого из шумов 60 дБ, то при сложении интенсивностей суммарный уровень будет 63 дБ, и громкость вырастет от 4 сон до 4,92 сона.

- если сигналы имеют разность частот шире критической полосы и их взаимным маскированием можно пренебречь, тогда действует другое правило: *суммарная громкость равна сумме громкостей каждой из составляющих*.

Отсюда получается, что при звучании сигналов с частотами, разнесенными шире критической полосы, суммарная громкость будет больше. Например, если два узкополосных шума имеют громкость по 4 сона, но частоты их разнесены (600 и 1200 Гц), то суммарная громкость будет 8 сон (а не 4,92 сона, как в предыдущем примере), что соответствует уровню громкости 70 фон.

- если частоты различных сигналов разнесены по частоте друг от друга достаточно далеко, то определение суммарной громкости значительно усложняется: слушатель обычно фокусирует свое внимание на каком-то одном компоненте (или самом громком, или одним из самых высоких), воспринимая общую громкость суммарного сигнала, примерно равной этому компоненту.

Для определения громкости сложного многокомпонентного звука в Международных рекомендациях ISO №532A рекомендуется использование следующей методики:

Рис. 6. Номограммы для расчета индексов громкости

с помощью стандартных октавных или третьоктавных анализаторов измеряется уровень звукового давления внутри каждой октавной (или третьоктавной) полосы. Затем с помощью графика (Рис. 6) определяется индекс громкости S_i (по оси ординат отложено значение центральной частоты), по оси абсцисс - значение уровня звукового давления в каждой полосе. Полученные значения индексов громкости S_i для каждой октавной полосы суммируются следующим образом:

$S = S_{\max} + 0,3(\sigma)S_i$, где S_{\max} - индекс самого громкого звука, $(\sigma)S_i$ - сумма индексов громкости во всех остальных полосах. Таким образом, суммарная громкость в сонах получается от суммирования 100% индекса громкости самого громкого звука и 30% от суммы всех остальных.

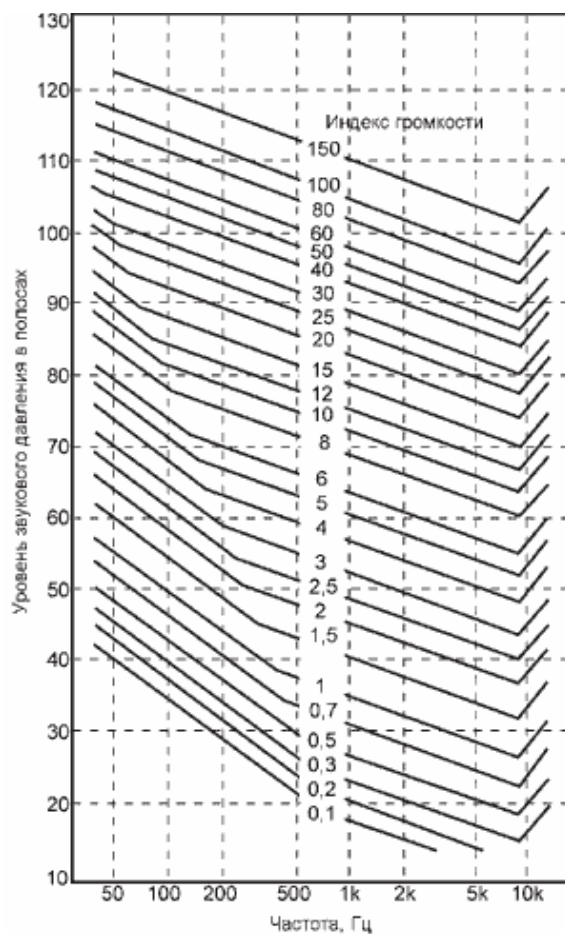
Значительно более сложный метод для оценки громкости реальных звуковых сигналов (шума, музыки и т.д.) был разработан Цвиккером (подробно изложен в книге "Ухо как приемник информации" Цвиккер Е., Фельдкеллер Р. Изд-во "Связь" М., 1973). Он позволяет оценить громкость комплексного сигнала с учетом взаимной маскировки его составляющих.

Этот метод введен в стандарты ISO532B и ANSI 3.4-1980. На его основе разработаны компьютерные методики расчета громкости и современные цифровые анализаторы громкости типа Zwicker Loudness Analysis Type 7704 со специальным программным обеспечением PULSE, что позволяет выполнять расчет громкости сложных стационарных сигналов в соответствии с международными стандартами, анализ громкости многоканальных нестационарных сигналов с учетом временных характеристик слуха, а также анализ спектрального распределения громкости и др. Подробное описание методики измерения и программного обеспечения можно посмотреть по адресу: <http://www.bk.dk/pulse/software///04.htm>.

Итак, ощущаемая громкость сложного звука зависит не только от его уровня интенсивности (уровня звукового давления), но и от его спектрального состава, что очень важно учитывать при создании музыкальных композиций. Например, звучание инструмента можно сделать более громким при сохранении того же уровня звукового давления за счет изменения его спектра (при этом, правда, произойдет и изменение тембра, так что все нужно делать в разумных пределах).

Интересно также посмотреть, как меняется уровень громкости при сложении основного и запаздывающего сигналов, что может привести (при прослушивании записей в сильно реверберирующем помещении) к существенному изменению баланса громкостей в звуковом материале.

При сложении основного и запаздывающего (например, отраженного) сигналов происходит приращение уровня громкости, при этом оно происходит по-разному в зависимости от общего уровня сигнала. Если сигнал



слабый (тихая речь - уровень громкости до 55 фон), то при величине задержки отраженного сигнала в 20...40 мс происходит увеличение уровня громкости на 3 фона, при дальнейшем увеличении времени задержки прирост уровня громкости снижается, т.к. сигналы начинают восприниматься раздельно (эхо). При уровнях громкости больше 55 фон увеличение общего уровня громкости происходит иначе - оно достигает 5 фон при задержке 50 мс, и затем также начинает снижаться.

В заключение приведу данные по уровням громкости (фон) и громкости (сон) для наиболее употребительных шумов и звуков, что может оказаться полезным для практической работы:

Источник звука или шума	Уровень громкости, фон	Громкость, сон
Шум в кабине самолета	128...130	875...1400
Средний шум на улице	55...60	3,08...4,35
Шум на улице с интенсивным движением транспорта	75...80	11,4...17,1
Звук оркестра	80...100	17,1...88
Шум аплодисментов	60...75	4,35...11,4
Разговор на расстоянии 1м:		
громкий	65...70	5,87...7,95
обычный	55...60	3,08...4,35
Шум в тихой комнате	25...30	0,2...0,36
Шепот на 1 м	20	0,1
Звук в радиостудии при исполнении соло	40...50	0,98...2,2
Шумное собрание	65...70	5,87...7,95

В музыкальной практике приняты, как известно, другие градации громкости. Их соответствие приведенным выше количественным оценкам громкости и уровням громкости приблизительно следующее:

Обозначение	Наименование	Уровень громкости,фон	Громкость, сон
fff	Форте-фортиссимо - самое громкое	100	88
ff	Фортиссимо - очень громкое	90	38
f	Форте - громкое	80	17,1
p	Пиано - тихое	50	2,2
pp	Пианиссимо - очень тихое	40	0,98
ppp	Пиано-пианиссимо - самое тихое	30	0,36

При создании компьютерных композиций, когда программно можно задавать большое число градаций громкости, полезно учесть, какие из них соответствуют значениям, принятым в музыкальной практике.

Часть 13. Субъективные критерии оценки акустики помещений.ч.1

В одном из предыдущих номеров ("Звукорежиссер", 7/2000) была опубликована статья, посвященная новой технологии создания виртуального трехмерного звукового мира - аурализации. В основе ее алгоритмов лежит обработка звукового сигнала, выполненная таким образом, чтобы вызвать у слушателя такое же ощущение восприятия музыки, как если бы он слушал ее в концертном зале, церкви, аудитории и другом помещении по его выбору.

Создание таких алгоритмов оказалось возможным только в настоящее время, и не только благодаря появлению новых возможностей компьютерной обработки звука, но и благодаря накопленному многолетнему опыту по выявлению основных критериев, которые определяют субъективное восприятие "хорошей" или "плохой" акустики зала. Прежде, чем переходить к новым разделам психоакустики: восприятию "тембра" и др., что планируется сделать в первых номерах следующего года, постараемся коротко рассмотреть основные критерии субъективной оценки акустики залов. Это важно не только для понимания новых технологий 3D-Sound, которые сейчас активно внедряются в практику работы со звуком, но и для работы с компьютерными аудиoproграммами, так как заложенные в них алгоритмы позволяют моделировать различные эффекты, имитирующие характеристики помещения. Звукорежиссеру необходимо понимать, к каким субъективным изменениям в восприятии звука может привести применение этих эффектов.

Любому музыканту, композитору, звукорежиссеру и просто любителю хорошей музыки прекрасно известно, какое огромное влияние оказывает на качество воспринимаемого звука акустика помещения, в котором исполняется музыка. Достаточно вспомнить, как звучит орган в огромном соборе, и представить, что

останется от этого звучания, например, в маленькой заглушенной комнате. Каждый стиль музыки требует своей оптимальной акустики зала, и композиторы прошлого учитывали это при создании своих произведений. Строительство хороших залов было и остается в значительной степени искусством, как и создание хороших музыкальных инструментов (скрипок, например), несмотря на огромные успехи, достигнутые в настоящее время в анализе объективных процессов формирования звукового поля в помещении.

Поскольку проблема расшифровки "слухового образа" остается еще окончательно не решенной, то и в оценке качества звучания в различных залах решающим является субъективная экспертиза. Поэтому за последние годы значительные усилия были приложены к установлению связи между объективно измеряемыми параметрами звукового поля в помещениях - и субъективной оценкой их качества звучания.

По этим вопросам в литературе опубликованы многочисленные, иногда противоречивые результаты. В качестве основы примем критерии, предложенные известнейшим специалистом в области акустики Беранеком.

Субъективная оценка акустики помещений для музыкальных и речевых программ представляет значительные трудности, поскольку требует решения следующих проблем: выбор метода оценки, выбор критериев оценки, установление их связей с объективными параметрами.

Выбор метода оценки

Все используемые методы представляют собой специально организованные тесты на прослушивание, которые проводятся тремя способами.

- слушатель производит оценку качества звука, находясь в синтезированном звуковом поле, создаваемом, например, распределенной системой громкоговорителей в заглушенной камере. (Рис. 1) Этот способ позволяет гибко менять и четко фиксировать параметры звукового поля: уровни звукового давления, время реверберации, время запаздывания и направление прихода ранних отражений и т.д. Такие эксперименты проводятся в достаточно большом объеме, особенно в Японии. Однако это трудоемкий эксперимент, кроме того, из-за конечного количества источников он создает упрощенную картину звукового поля в помещении;

- непосредственное прослушивание оркестра или исполнителей в испытываемых залах опытными экспертами с последующей статистической обработкой их оценок. (Рис. 2) Это наиболее точный метод, однако требует большого объема экспериментов, при которых трудно добиться повторяемости результатов, и сложно менять отдельные параметры;

- на основе стереофонических записей, сделанных в испытываемых залах с помощью "искусственной головы" (рис.3) и последующем прослушивании через головные телефоны или громкоговорители. Этот способ позволяет получить достаточно точные результаты, хотя техника бинауральной записи как таковая имеет свои проблемы. Такие эксперименты многократно проводились (прежде всего, в Германии) и были получены очень ценные результаты.

В любом случае, результаты субъективных оценок акустики помещений существенно зависят от выбора экспертов: их профессии, опыта прослушивания, вкусов общей и музыкальной культуры, и т.д.

Выбор критериев оценки

Одной из первых попыток установить "словарь" критериев субъективной оценки акустики музыкальных залов была предпринята Беранеком. На основе личного опыта, а также из бесед с известными дирижерами, музыкантами, опытными слушателями, он выбрал из многочисленных субъективных оценок различных залов (теплый, холодный, пустой, глухой и др.) восемнадцать наиболее употребляемых субъективных критериев, а из них десять наиболее значимых и независимых. Хотя эта методика вызвала ряд возражений специалистов, но она послужила толчком к многочисленным исследованиям, и в настоящее время некоторые из результатов этих исследований введены в стандарты.



Рис 1 Моделирование параметров помещения в заглушенной камере

К наиболее распространенным субъективным критериям для оценки акустического качества помещений относятся: гулкость, жизненность (liveness); полнота звука (fullness); различимость или ясность (definition или clarity); интимность (intimacy), теплота (warmth), пространственность (spaciousness), громкость (loudness); баланс (balance); ансамбль (ensemble); тембр (timbre), а также отрицательные факторы: эхо, порхающее эхо (flutter), мешающие шумы.

Установление связей объективных параметров качества звучания и их субъективных оценок

Прежде чем приступить к решению этой задачи, была выполнена большая работа по общей классификации всемирно известных концертных залов по качеству звучания в них различных музыкальных произведений на основе анкетных опросов музыкантов, музыкальных критиков, опытных слушателей и т.д. В результате все рассмотренные залы (а было изучено более пятидесяти известных залов в разных странах мира), были сгруппированы в три группы - А, В, С в соответствии с качеством звука исполняемых в них произведений.

Соответственно, в этих залах были проведены измерения объективных параметров реверберационного процесса. Вопрос о выборе наиболее значимых объективных параметров, современных компьютерных методах расчета и измерения структуры звукового поля и временной структуры процессов затухания звука (т.е. параметров реверберационного процесса) в помещениях различной конфигурации заслуживает отдельного разговора. Здесь коротко остановимся только на некоторых параметрах, которые были использованы Беранеком в процессе анализа субъективных оценок.

Звук, который достигает слушателя в любом помещении прослушивания, содержит информацию как о параметрах звука, созданных музыкальным инструментом, певцом и т.п., так и о свойствах помещения, в котором этот звук воспроизводится. Помещение прослушивания (студия, концертный зал, стадион и др) является своего рода линейным фильтром, который производит обработку поступившего в него звукового сигнала, изменяя его временную структуру и изменяя его спектр, что, соответственно, приводит к изменению его тембра и определяет качество звучания.

Обусловлено это, прежде всего, тем, что в помещении, наряду с прямым звуком, к слушателю приходят многочисленные отражения, которые и формируют структуру реверберационного процесса, характерную для каждого вида помещения - она зависит от его размера, формы, отделки, наличия слушателей и др. Пример реверберационного процесса показан на рисунке 4. Как видно из рисунка, в начальный момент при использовании в качестве источника, например, короткого импульсного сигнала, к слушателю поступает прямой звук, затем, через определенное время, начинают поступать отражения от различных поверхностей, которые сначала четко разделены друг от друга по времени, затем количество их увеличивается, звуковое поле становится диффузным, и уровень звукового давления в данной точке помещения постепенно спадает - такой процесс спада звука в помещении и называется реверберационным.

Для описания параметров реверберационного процесса обычно используется величина времени реверберации (ВР), которая определяется как "время, в течение которого уровень звукового давления уменьшается на 60 дБ". Величина времени реверберации определяется объемом зала и общим коэффициентом звукопоглощения в нем, она не зависит от формы зала, структуры распределения поглощающего материала и т.д., т.е. является усредненной характеристикой.

Однако исследования по оценке качества звучания в различных помещениях заставили ввести целый ряд дополнительных параметров, более тонко характеризующих реверберационный процесс.



Рис 2 Концертный зал

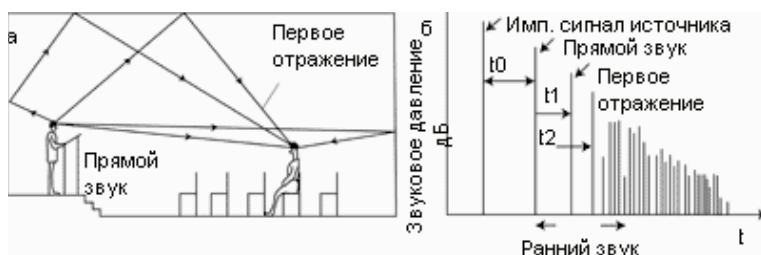


Рис 4 Структура реверберационного процесса

"Ранний звук" определяется как прямой звук и отражения, поступившие в течение первых 80 мс после прихода прямого звука. Причем существенное значение имеет направление прихода этих ранних звуков - так, например, звуки, пришедшие от боковых стен в первые 80 мс, создают ощущения расширения источника звука, что улучшает качество восприятия музыки.

"Громкость ранних звуков" определяется энергией прямого звука плюс энергия отраженных звуков, пришедших в первые 80 мс. "Громкость реверберирующего звука" определяется общей звуковой энергией, которая достигает слушателя после 80 мс.

"Раннее время реверберации" (РВЗ) - время затухания звука после выключения источника, когда уровень звукового давления уменьшается на 10 дБ. Кроме того, для сопоставления с субъективными оценками используется также время реверберации (ВР) при спаде звукового давления от -5 до -35 дБ при заполненных залах (т.е. часть реверберационной кривой).

"Коэффициент внутрислуховой кросс-корреляции (КВСКК)" определяется как коэффициент корреляции сигналов, поступивших на два уха при разном времени интеграции и в разных частотных диапазонах. Обычно используется время интеграции от 0 до 80 мс в трех октавных полосах: 500, 1000 и 2000 Гц. Этот коэффициент характеризует степень различия звуковых сигналов в двух ушах как по времени, так и по амплитуде.

Кроме этих, для сравнения с субъективными оценками используется и целый ряд других параметров: эквивалентная реверберация, распределение уровней звукового давления и др.

Сравнение результатов субъективных экспертиз, проведенных в вышеуказанных залах, с приведенными выше параметрами, показало отчетливую связь между общим впечатлением от акустики зала и временем реверберации, причем в качестве времени реверберации ВР использовалось значение времени затухания от уровня -5 до -35 дБ (т.е. часть кривой затухания) для заполненных публикой залов, и время ранней реверберации РВЗ от 0 до -10 дБ. Как следует из таблицы, при переходе от залов группы А к группе В и С величины ВР и РВЗ изменяются от среднего значения 2 и 2,6 с (группа А) до 1,6 и 1,9 с (группа В) и до 1,4 и 1,75 с (группа С).

На основе полученных результатов была выполнена более дифференцированная оценка отдельных субъективных параметров акустики залов, и исследована их связь с измеренными объективными характеристиками.

"Гулкость-жизненность" - эти термины прежде всего связаны с оценкой общего впечатления от акустики залов, т.е., как было указано выше, в значительной степени связаны с временем реверберации (ВР). Все помещения по этому критерию ранжируются достаточно четко (соборы, концертные залы, студии и др.). Для каждого вида музыки и речи существуют оптимальные пределы изменения времени реверберации, которые зависят от объема помещения и частоты (например, рис. 5), которое меняется в пределах от 0,4 до 1 с для речи, от 1 до 1,5 с для камерной музыки, от 1,8 до 2,2 с для симфонической и т.д. Реверберация - один из эффектов, который учитывался композиторами при создании произведений, например, композиторы органной музыки специально делали паузы, чтобы была слышна длинная реверберация в соборе.

Наибольшее влияние на ощущение "жизненности" звуков оказывает значение времени реверберации на средних частотах. В помещениях, в том числе в студиях, где время реверберации слишком короткое для данного музыкального жанра, звук характеризуется как "мертвый", "сухой". Все, кто слышал, как звучит музыка в заглушенной камере, могут отчетливо представить, что именно имеется в виду. Наоборот, если время реверберации слишком велико для данной музыки, звук характеризуется как слишком "грязный", "водянистый". Поэтому введение при обработке дополнительных эффектов реверберации, не соответствующих стилю и характеру музыки, может вызвать аналогичные субъективные ощущения.

Полнота тона (звучность) также зависит от времени реверберации, но также и от отношения громкости реверберирующих звуков, которая определяется энергией звуков, приходящих после первых 80 мс (E_1) к громкости ранних звуков, которые определяются энергией прямого звука и первых отражений до 80 мс (E_2).

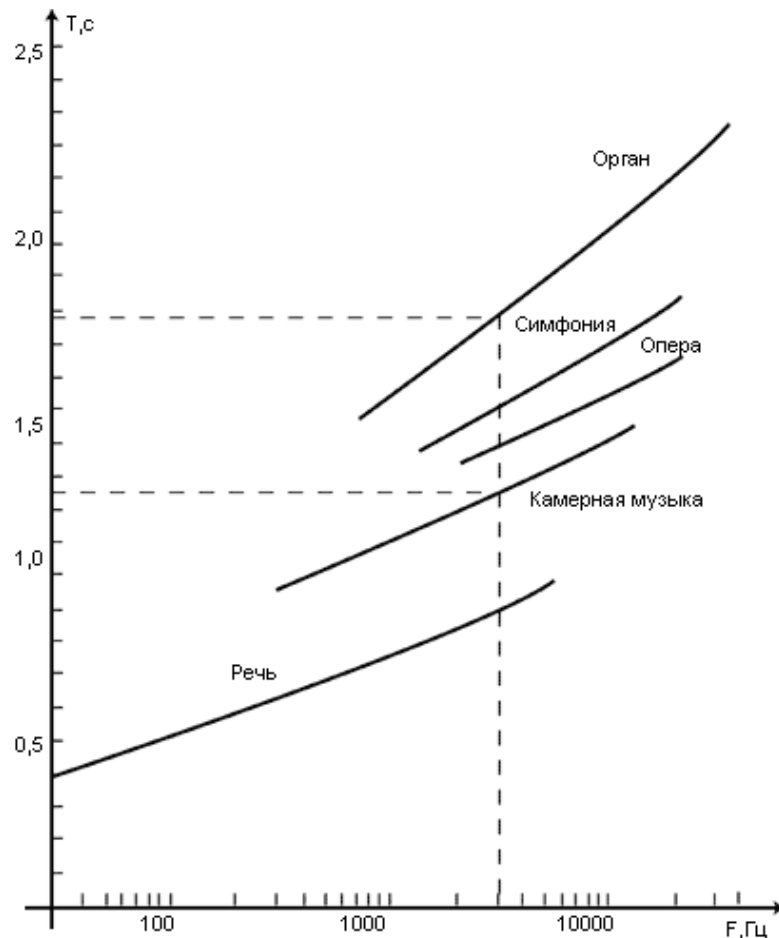


Рис 5 Оптимальное время реверберации

Чем больше отношение $E1/ E2$, тем выше "полнота тона".

Для церквей это отношение велико, и звук воспринимается как "полнозвучный". В помещениях, где энергия отраженных звуков мала, звук будет казаться "пустым". В залах (например, старинных оперных театрах), где звук от исполнителя имеет возможность свободно подниматься и отражаться от высоких потолков, энергия в реверберирующих звуках будет достаточно большой и звучание также будет "полным". Для обеспечения этого качества звука большое значение имеет выбор формы зала и размещение специальных отражающих панелей и других деталей убранства.

В опциях reverb почти всех компьютерных аудиoproграмм пользователю предоставляется возможность произвольно менять соотношение ранней и поздней частей энергии в реверберирующем звуке. Однако использование этой возможности без понимания того, к каким изменениям субъективного восприятия это можно привести, может сделать обрабатываемую композицию "сухой" и "пустой". Допустимые пределы изменения этого параметра для хороших залов будут приведены ниже.

Различимость и ясность. Когда музыканты говорят о "различимости" или "ясности", имеется в виду степень, с которой отдельные звуки в музыкальном произведении четко разделяются друг от друга. Имеется два вида "ясности" (различимости): "горизонтальная" и "вертикальная".

Горизонтальная относится к звукам, следующим друг за другом. Композитор использует специальные приемы, чтобы обеспечить ее: темп, повторение тонов во фразе, относительную громкость последовательных тонов и т.д. Исполнитель также может влиять на горизонтальную различимость выбором манеры исполнения.

Акустические факторы в помещении, которые определяют "горизонтальную различимость" музыкального произведения - это величина времени реверберации и отношение громкости (энергии) ранних звуков к громкости (энергии) реверберирующего звука: $C80= E2/E1$, т.е. факторы те же, но отношения обратные. Таким образом, увеличение "горизонтальной различимости" уменьшает полноту тона, и наоборот.

Список концертных залов, разделенных по категориям акустического качества

Концертные залы	Число мест	Объем, м3	ВР для заполненных залов, с	РВЗ для заполненных залов, с
Группа А:				
Бостон, Симфонический зал	2625	18750	1.8	2.4
Вена, Гроссер Мюзикверейнс Зал	1680	15000	2	3
Амстердам, Консерт гебоув	2047	18780	2	2.6
<i>Средние значения</i>	2047	18750	2	2.6
Группа В:				
Чикаго, Оркестр Холл	2582	15180	1.3	
Сан-Франциско, Дэвис Холл	2843	24350	2.2	2.2
Эдмонт, Альберта Джабили Холл	2731	21500	1.4	1.4
Монреаль, Салле Уилфред Пеллетье	2998	25000	1.7	1.9
Ванкувер, Театр королевы Елизаветы	2800	21500	1.5	
Тель-Авив, Манн Аудиториум	2715	21240	1.6	1.7
<i>Средние значения</i>	2760	21500	1.6	1.9
Группа С:				
Буффало,Клейнханс Мюзик Холл	2839	18240	1.3	1.6
Блумингтон, Зал Университета	3788	28700	2.2	
Лондон, Барбикан Консерт Холл	2026	17750	1.7	1.9

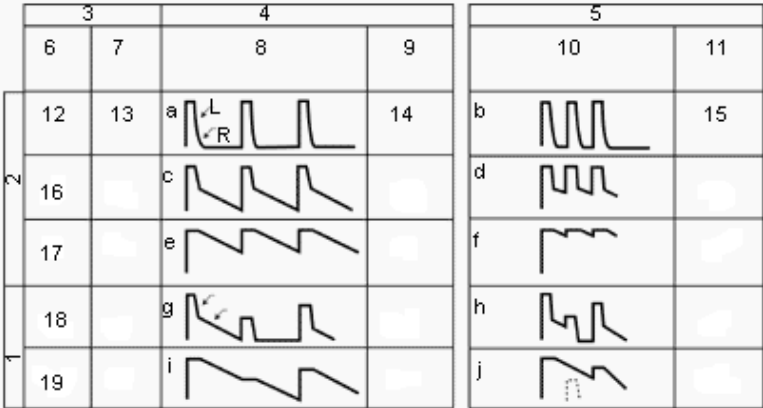


Рис 6 Таблица для оценки параметров помещения
1 - Звуки разн громкости 2 - Звуки одинак громкости 3 - Акуст условия 4- Звуки исп медленно 5 - Звуки исполняются быстро 6 - Время реверберации (б - большое м - малое с - среднее) 7 - C80 8 - Музыкально-акуст результаты 9 - ясность/полнота (далее я/п, в - высокая, с - средняя, н - низкая) 10 - Музыкально-акуст результаты 11- ясность/полнота 12 - м 13 - б 14 в я/п 15 в я/п (следующая сттрока) - б с в я/сп в я/сп 17 - б.м.с/я/п.н/я/п 18- б.с.в я/сп. в я/сп 19 б.м.н/я/п очень н/я/п

Вертикальная различимость - это степень, с которой звуки, звучащие одновременно, различаются на слух. Она также зависит от стиля произведения, искусства исполнителя, акустики зала и тренированности слуха.

Композитор влияет на нее выбором одновременно звучащих тонов, выбором инструментов и т.п., а исполнитель может влиять, меняя динамику звучания одновременных тонов и др.

Акустические факторы для вертикальной различимости - баланс звуков различных инструментов, который существенно зависит от акустических параметров сценического пространства; и также отношение энергии ранних звуков к энергии реверберирующего звука.

Таким образом, горизонтальная и вертикальная различимость (ясность) зависит как от музыкальных, так и от акустических факторов. Их влияние должно быть заложено в замысел композитора, чтобы сделать музыку доступной аудитории. Например, органные хоралы Баха, с их крупными длительностями, медленными мелодическими линиями и растянутой динамикой, требуют помещений с большим временем реверберации (более трех секунд) и высоким значением отношения полной реверберирующей энергии к энергии ранних отражений. Они имеют малую горизонтальную различимость, но высокую полноту тона.

Концертам Моцарта, с быстрыми пассажами фортепьяно и развитой оркестровой фактурой, в противоположность органной музыке, необходимы помещения с относительно коротким временем реверберации и большим отношением ранней к реверберирующей энергии, т.е. с высоким горизонтальным и вертикальным разрешением.

Влияние темпа исполнения музыки на ощущения полноты и ясности звучания в помещениях с разными значениями времени реверберации и разным отношением $E2/E1$ показано в таблице на рисунке 6. Здесь показана разделимость отдельных коротких звуков в музыкальном произведении в зависимости от времени реверберации BP и ясности $S80$. Как видно из таблицы, для помещений с коротким временем реверберации и большим $S80$ (первая строка, примеры а и б) индивидуальные тоны быстрой и медленной музыки разделяются отчетливо, и процессы их атаки и спада хорошо различимы, заметна быстрая часть спада звука самого инструмента I , и более медленная часть R из-за процесса реверберации.

Во второй строке, что соответствует помещениям с большим временем реверберации и средним значением ясности $S80$ (примеры с и d) атака отдельных звуков будет слышна, а часть участка спада звука инструмента будет "закрита" реверберацией (т.е. звучание инструмента существенно "затягивается" за счет того, что звук в помещении затухает достаточно медленно). При более быстром темпе исполнения уже и часть атаки, так же как и спада звучания инструмента, будут плохо различимы из-за процесса реверберации.

Для примера е и f длина BP та же, но отношение раннего звука к реверберирующему звуку меньше, при этом часть атаки и почти весь спад звука инструмента "закрываются" реверберацией, тоны различимы плохо, но полнота звука большая. При быстром темпе f тоны едва различимы, они почти полностью скрыты реверберацией. Полнота звуков максимальная, но играть staccato в таком помещении невозможно. Еще больше влияние реверберации и ясности $S80$ сказывается на исполнении звуков разной громкости: примеры g и h и примеры i и j. Как видно из таблицы, слабые тоны практически полностью маскируются процессом реверберации и становятся неразличимыми.

Эти соотношения в конкретном зале должны иметь в виду композитор и исполнитель, выбирая темп, фразировку и т.д.

Таким образом, разные стили музыки требуют различных значений вышеуказанных параметров. О величине времени реверберации для концертных залов и музыкальных студий было сказано выше, что касается коэффициента ясности (различимости) $S80$, то для залов, оцененных музыкантами-экспертами как залы с хорошей различимостью, его значения находятся в пределах от -3,7 до -0,02 (среднее значение -2,5).

Современные компьютерные технологии дают возможность менять параметры, моделирующие процессы реверберации в разных помещениях, в очень широких пределах. Достаточно сложные алгоритмы, например Acoustic Modeller, позволяют осуществлять "свертку" сигнала с импульсными характеристиками различных помещений, что дает возможность заставить звучать музыкальную композицию так, как если бы она звучала в этих залах. Однако, выбирая параметры этих помещений или создавая собственные, всегда необходимо помнить, что несоответствие характеристик помещения (времени реверберации, отношения энергии ранних звуков к поздним и др.) стилю музыкального произведения и темпу его исполнения, может привести к совершенно противоположным ощущениям, чем это предполагалось, поэтому при создании электронных композиций или обработке фонограмм необходимо учитывать указанные критерии.

Часть 13. Субъективные критерии оценки акустики помещений, часть 2

В первой части статьи ("Звукорежиссер", 10/2000) было отмечено, что в результате многочисленных экспертиз, выполненных известнейшим специалистом-акустиком Беранеком, было установлено, что к наиболее

распространенным субъективным критериям оценки акустического качества помещений относятся: гулкость, жизненность (liveness), полнота звука (fullness), различимость (definition) или ясность (clarity), интимность (intimacy), теплота (warmth), пространственность (spaciousness), громкость (loudness), баланс (balance), ансамбль (ensemble), тембр (timbre), а также отрицательные факторы: эхо, порхающее эхо, мешающие шумы.

Была рассмотрена связь таких субъективных критериев, как гулкость, жизненность, полнота звука, различимость или ясность с объективными параметрами, характеризующими реверберационный процесс в помещении: время реверберации, отношение энергии ранних и поздних отражений и др.

Рассмотрим следующие критерии: интимность (присутствие, камерность, близость). Они определяют для слушателя кажущийся размер пространства, в котором он слушает музыку. Разные стили музыки требуют разных значений "акустической интимности". Интимность определяется разницей во времени между прямым и первым отраженным звуками, а также, частично, общей воспринимаемой громкостью звучания, так как слушатель предполагает, что звук в маленьком помещении кажется громче, чем в большом. Основной вклад в ощущение "интимности" вносят первые отражения от боковых стен (в залах с достаточно высокими потолками), или от потолков при их сравнительно низкой высоте.

Рис. 1. Разница путей прямого и отраженного звукового луча в залах различных размеров

Как видно из рисунка 1, разница путей прямого (D) и отраженного (R1) звуков в первом зале меньше, чем во втором, и, соответственно, различается разница во времени прихода звуков, равная $t_n = (R_n - D)/C$, где C скорость звука, а n = 1,2. Это, естественно, приведет к тому, что интервал времени между прямым звуком и первым отражением (рис. 2) во втором зале будет больше. Большая разница во времени прихода прямого звука и первого отражения создает у слушателя ощущение оторванности (отдаленности) от исполняемой музыки.

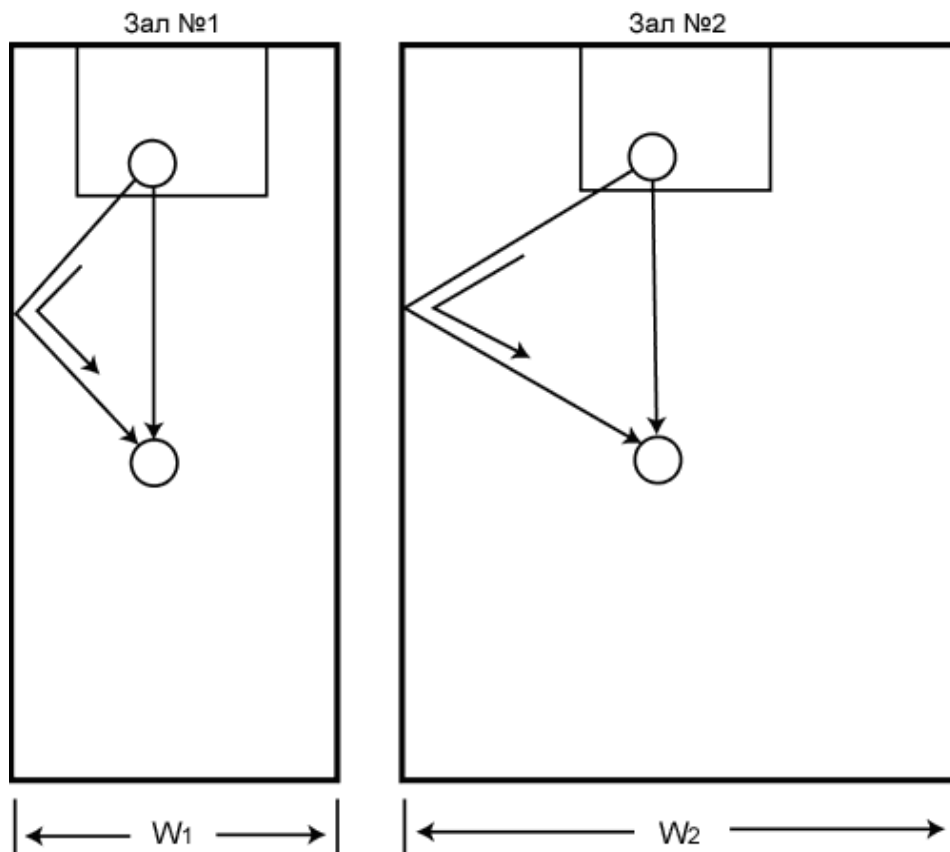


Рис. 2. Интервал времени между прямым звуком и первым отражением

В залах с хорошей акустикой, предназначенных для симфонического репертуара, этот временной интервал составляет для слушателей, сидящих в центре зала, величину 15-30 мс. В пределах этого времени, если отражения имеют похожий спектр и огибающую, а их громкость не выше прямого звука, они не воспринимаются как отдельные звуки, а помогают в улучшении локализации прямого звука (это явление известно как эффект Хааса). Для скорости звука 340 м/с задержка на это время соответствует разнице в расстоянии примерно 12 м, что требует ширины зала порядка 18-20 м. Композитор (звукорежиссер, исполнитель и др.) должен иметь в виду этот параметр, иначе будет несоответствие размеров помещения стилю музыки, которое очень четко ощущается слушателями, примером может служить звучание органа в маленькой комнате.



В старинной музыке (до 17 века) камерные произведения создавались в основном для малых ансамблей, и

исполнялись в залах с малой разницей прихода ранних отражений (меньше 15 мс), что создавало ощущение близости ("интимности") звучания. В 18-19 веках изменился стиль музыки, увеличились исполнительские составы и размеры помещений (оперные театры, концертные залы и др) и, соответственно, выросло время задержки до 30 мс. В настоящее время, когда многие концертные залы имеют очень большие размеры, исполнение в них камерной музыки создает ощущение несоответствия размеров зала стилю. Музыка как бы теряется в зале. Для улучшения этой ситуации иногда используются дополнительные отражающие поверхности около сцены по бокам или на потолке, что позволяет создать дополнительные ранние отражения с меньшим временем запаздывания, и тем самым улучшить восприятие исполняемой музыки.

Пространственность - ощущение слушателя, что музыка идет от полной ширины зала, и звук окружает его со всех сторон, что обычно характеризует залы с хорошей акустикой. Наиболее полно это ощущение проявляется при прослушивании, например, органа или хора в больших соборах. В противоположность этому, в залах с плохой акустикой звук кажется идущим как бы из "окна".

Тренированный слушатель (тем более звукорежиссер) может различить две составляющие в восприятии пространственности кажущееся расширение площади источника звука (ASW) и окружение (или "обертывание" LEV), когда слушатель чувствует себя погруженным в звук со всех сторон.

По мнению многих экспертов, первая составляющая является одним из главных индикаторов акустического качества концертных залов и помещений прослушивания. Она связана с уровнем боковых ранних отражений: чем выше уровень боковых отражений в помещении, тем больше кажущееся расширение источника. Кажущаяся ширина звукового источника связана также с уровнем громкости на низких частотах C_{3n2} (в основном в области частот 125 и 200 Гц).

Однако наибольшую связь с этим параметром показали результаты измерения коэффициента внутрислуховой кросс-корреляции сигнала KBCKKp3. Этот коэффициент определяет степень разности звуковых сигналов в двух ушах как по времени, так и по амплитуде. Чем менее сходны звуки в левом и правом ушах, тем меньше этот коэффициент, и тем больше кажущееся расширение источника $ASW=1-KBCKKp3$. В случае, если звуки одинаковы, коэффициент становится равным единице, и кажущийся источник звука концентрируется в центре. Эти результаты подробно разработаны в теории стереофонии.

Измерения, выполненные в различных залах с помощью прибора "искусственная голова" на двадцати позициях при разных положениях источников звука, показали, что значения этого коэффициента, усредненного в трех октавных полосах 500 Гц, 1 и 2 кГц, при интеграции по времени в интервале 0-80 мс KBCKKp3 (p-ранний по времени прихода, 3-усредненный в трех полосах), дают хорошую корреляцию с субъективными оценками кажущегося расширения источника - ASW.

Все измеренные залы оказались четко ранжированы по этому параметру: для лучших по качеству звучания залов мира значения этого коэффициента KBCKKp3 оказались в пределах 0,3-0,6.

Обертывание (погружение LEV) связано с ощущением позднего реверберирующего звука, поступающего со всех сторон (после 80 мс), и зависит от конструкции зала: наличия нерегулярностей стен, балконов и т.д., т.е. всех конструктивных элементов, которые обеспечивают приход звука с разных сторон (диффузность звукового поля). Так, например, ощущения звучания музыки у слушателя, к которому отраженные звуки приходят со всех сторон: от потолка, стен, пола и т.д., будут существенно отличаться от ощущений сидящего под балконом слушателя, к которому звук приходит только с фронта. Значение этого коэффициента также связано с коэффициентом внутрислуховой кросс-корреляции, усредненным за период времени от 80 мс до 1 с, однако статистически значимых измерительных данных по этому параметру еще не набрано.

Громкость - для оценки залов используется специальное субъективное понятие, характеризующее громкость источника звука при игре фортиссимо по отношению к некоторой "ожидаемой" громкости на месте прослушивания. Наиболее благоприятное расстояние по этому параметру в большинстве залов для прослушивания прямого звука от оркестра 18 м, от солистов 6-15 м.

Громкость определяется субъективным ощущением силы звука, она пропорциональна плотности звуковой энергии на месте прослушивания.

Для объективной оценки громкости предложен такой параметр как сила звука C_3 , который определяется как разность уровней звукового давления, измеренного на шумовом сигнале на месте слушателя, и уровнем звукового давления от того же источника на том же расстоянии в заглушенной камере, при этом измерения проводятся в октавных полосах с частотами 125, 250, 500 Гц, 1, 2 и 4 кГц. При измерениях учитывается только энергия прямого звука и ранних отражений, пришедших в первые 80 мс. Обычно нормируются два параметра: один, усредненный в полосах 125 и 250 Гц (C_{n2}), и другой в полосах 500, 1000 и 2000 Гц (C_{p3}). Значения этих параметров для лучших залов оказались равными 6 и 6,2.

Теплота - отношение времени реверберации на низких частотах к времени реверберации на средних. Оно измеряется при заполненном зале, при этом время реверберации на низких частотах (125 Гц) должно быть равно времени реверберации на средних частотах (500-1000 Гц), или быть больше примерно на 20%. "Теплота" субъективно определяется как звучность басов по сравнению со звучностью средних частот. Беранеком был предложен критерий КНТ коэффициент низкого тона, равный отношению среднего значения

времени реверберации ВР на частотах 125 и 250 Гц к среднему значению времени реверберации на частотах 500 и 1000 Гц:

$$KTH = \frac{BP_{125} + BP_{250}}{BP_{500} + BP_{1000}}$$

Его измеренные значения для лучших концертных залов оказались в пределах 1,08 1,1.

Баланс - понятие, служащее для оценки громкости отдельных инструментов и групп инструментов по отношению к общей громкости оркестра. Баланс должен быть как между разными группами инструментов, так и между оркестром и солистами. Баланс зависит от особенностей околосценического пространства, размещения оркестрантов, от исполнительской концепции дирижера, звукорежиссерского решения, и др.

Ансамбль - понятие, включающее в себя стройность, слаженность совместного исполнения, в том числе ритмическую точность и синхронность исполнения отдельных партий. Чувство ансамбля зависит от слышимости собственного исполнения и взаимной слышимости, что определяется в значительной степени конструкцией сцены и поверхностей вблизи нее.

Вопросами акустики сцены и околосценического пространства, и их влияния на общее звуковое впечатление в зале, сейчас занимается много исследователей, и получены достаточно интересные результаты.

Тембр - понятие сложное и многогранное, проблемам его восприятия уделяется сейчас особое внимание в психоакустике. Этой теме будут посвящены очередные статьи. Можно пока принять определение, предложенное Беранеком, хотя имеются и другие: "тембр качество звука, иногда его называют "окраской звука", которое позволяет отличить звук одного инструмента или голоса от другого". Каждый инструмент имеет свой характерный тембр звучания, достаточно вспомнить исполнение одной и той же мелодии на разных инструментах, например, на фортепьяно или органе, который определяется его конструкцией и материалами, из которых он изготовлен.

Рис.3. Осциллограмма звучания звука скрипки C4

Тембр зависит от структуры звука во все периоды его звучания: в момент установления, в стационарный период и в момент спада. Осциллограмма звука скрипки C4 и трехмерный спектр показаны на рисунках 3 и 4. Акустические свойства помещения оказывают влияние на все этапы звучания и, соответственно, на воспринимаемый тембр. Как уже было отмечено в первой части этой статьи, помещение является линейным фильтром, который производит обработку музыкального или речевого сигнала как во временной, так и в частотной областях. Реверберационный процесс в помещении изменяет характер процессов нарастания (атаки) и спада звука, при этом структура распределения резонансных частот в помещении существенно влияет на его спектр. Для помещений прямоугольной формы с отражающими стенами резонансные частоты могут быть рассчитаны по простой формуле:

$$f_{k,m,n} = c/2 [(k/L)^2+(m/B)^2+(n/H)^2]^{1/2},$$

где c - скорость звука, f k,m,n - частоты резонансных колебаний, L, B, H длина, ширина и высота помещения, k, m, n целые числа, определяющие номер моды (формы) колебаний.

Рис.4. Спектр звука C4

Как видно из этой формулы, значения резонансных частот зависят от размеров помещения. Результаты расчетов резонансных частот для больших помещений и для помещения с малыми размерами показывают, что первые дискретные резонансы для помещения с малыми размерами попадают в область слышимых частот, и значительно окрашивают звучание.

На рисунке 5 показан пример спектра для помещения с размерами 9 x 7,5 x 5,8 м.

Рис.5

На рисунке 6 - пример неравномерного распределения звукового давления в помещении на первой осевой резонансной частоте.

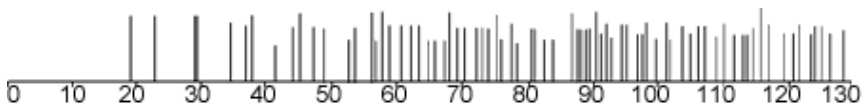
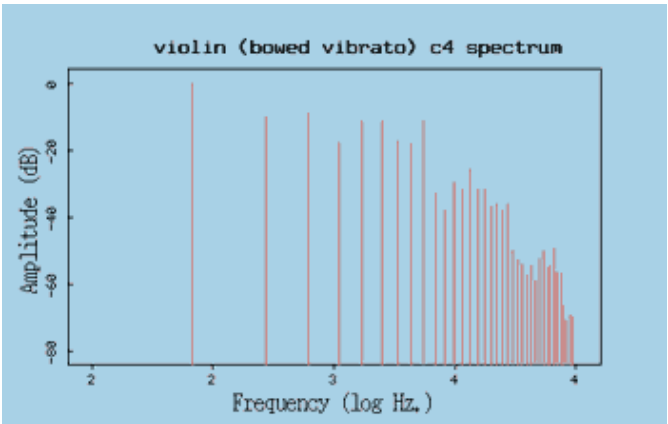
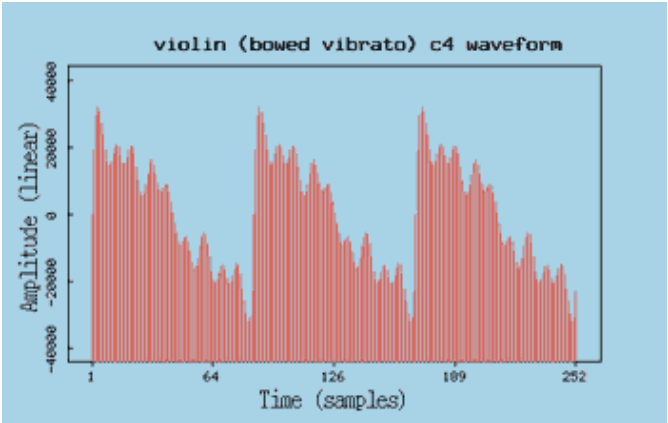
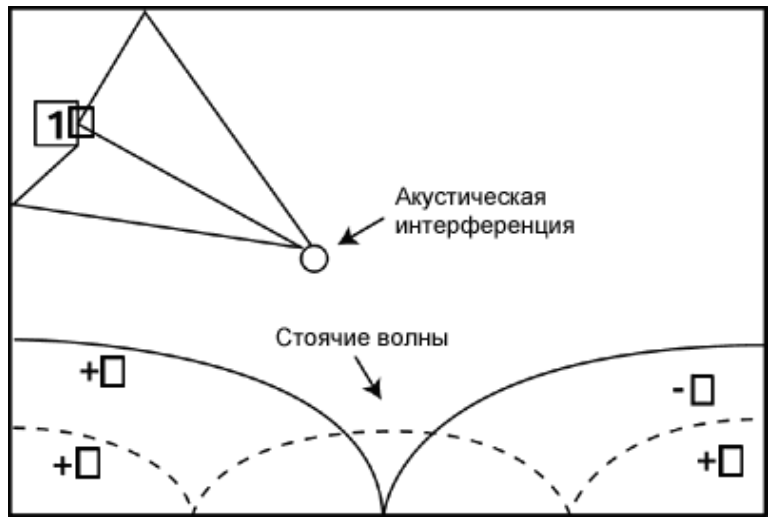


Рис.6



Плотность резонансных частот должна быть достаточно велика, чтобы не было заметно изменение тембра за счет резонансов. Исполнение музыки в помещениях с малым объемом неизбежно приводит к искажению тембра за счет дискретных резонансов, малого времени реверберации, недостаточного временного интервала между прямым звуком и первыми отражениями, и др. Музыка требует пространства, в частности, для музыкальных студий минимально допустимый объем должен быть не менее 200 м

Таким образом, тембр существенно зависит от размеров и формы помещения, от распределения и величины затухания в нем, от наличия рассеивающих элементов, обеспечивающих диффузное звуковое поле, и многих других факторов. С учетом всего этого и складывается искусство создания залов с хорошей акустикой. Следует подчеркнуть, что влияние параметров помещения на тембр звучания очень существенно, о чем по собственному опыту хорошо знают музыканты-исполнители. Поэтому выбор оптимального значения времени реверберации для каждого стиля музыки, а также оптимальных значений рассмотренных выше других параметров, является обязательным условием обеспечения хорошего звучания музыкальных и вокальных произведений в данном помещении.

Созданные за последние годы методы аурализации компьютерного моделирования звукового поля в помещениях, дающие возможность предварительного прослушивания звучаний различных источников (см. статью "Аурализация виртуальный звуковой мир", журнал "Звукорежиссер" 7/2000) открывают принципиальные новые возможности в проектировании концертных залов, студий и др., позволяя проверить различные варианты акустических решений и оценить их влияние на тембр звучания музыки и речи. На это раньше уходили годы, и поиск проходил методом "проб и ошибок".

Наряду с акустическими параметрами помещений, которые определяют положительные впечатления от прослушивания в них музыки и речи, существует целый ряд параметров, которые являются мешающими факторами при прослушивании. К числу основных негативных факторов относятся:

Общий вид музыкальной студии Royaltone Studio

Эхо - заметное на слух повторение прямого звука. Заметность эха зависит от времени запаздывания и интенсивности отраженных сигналов. При времени запаздывания меньше 80 мс ощущение эха практически отсутствует даже при достаточно больших уровнях сигнала.

Порхающее эхо - многократная периодическая последовательность эха, что создает тональную окраску звука (эффект гребенчатого фильтра), особенно если период последовательности меньше 20 мс. Сильный эффект наблюдается при наличии длинных параллельных стен, что характерно для многих современных залов. Присутствие эха может приводить к нарушению локализации звуковых источников, что совершенно недопустимо в помещениях для прослушивания музыки.



В классических залах использовались специальные рассеивающие поверхности: колонны, ложи и др., а также выбиралась не прямоугольная форма помещения, что и сейчас используется в хороших "акустических" студиях (см. фото).

Мешающие шумы - общее впечатление от любого исполнения музыки или речи может быть в значительной степени испорчено, если в зале имеется высокий уровень мешающих внешних или внутренних шумов. В таких залах или студиях оказывается трудным, иногда практически невозможным, обеспечить звукозапись, и даже очистить фонограммы от шумов различного происхождения оказывается не всегда возможным, несмотря на наличие современных компьютерных технологий.

Уровень шумов в зале определяет динамический диапазон воспринимаемого музыкального или речевого сигнала, поскольку слабые уровни сигнала маскируются шумом, что приводит к значительной потере качества звучания музыки или к потере разборчивости речи.

Уровень шумов в помещении зависит от нескольких причин:

-проникновение внешних шумов от транспорта и др. Именно для борьбы с этими шумами применяются различные способы звуко- и виброизоляции при строительстве концертных залов и студий. Для этого строят студии в тихих местах, используют дополнительные стены на отдельном фундаменте типа "здание в здании",

применяют специальные звуко- и виброизоляционные материалы и т.д.;
 -возникновение внутренних шумов от вентиляционных, осветительных и других систем, а также шума от публики. При строительстве студий затрачиваются значительные средства на уменьшение уровней шумов от различных обеспечивающих систем.

Рис.7. Нормирующие кривые уровня шумов

Общий уровень шумов в хороших залах и студиях должен соответствовать международным нормам, т.е. быть ниже кривой NC-20, предпочтительнее NC-15. Соответствующие кривые допустимых шумов показаны на рисунке 7.

Общее акустическое впечатление от помещения, в котором прослушиваются музыка и речь, складывается из всех вышеперечисленных факторов.

Значения основных вышеуказанных параметров для трех залов, отнесенных экспертами к лучшим залам мира по качеству звучания музыки, приведены в таблице 1:

Название зала	Время реверберации ВР, с	Кол-во мест	Коэф. кросс-корреляции 1-КВСККр3	Ясность С80, дБ	Громкость СЗр3, дБ	Громкость на НЧ, СЗнч2, дБ	Разность времени прихода t1, мс
Мьюзик Верейнс-Саал, Вена	2,0	1680	0,71	-3,7	6,8	6,9	12
Симфонич. зал, Бостон	1,8	2625	0,65	-2,7	4,8	2,6	15
Концерт-Гебоув, Амстердам	2,0	2047	0,62	-3,3	5.8	5,1	21

Классификация лучших залов мира ,выполненная в работах Беранека по результатам измерений и субъективной оценки более 50 концертных залов, показала, что значения указанных выше объективных параметров существенно меняются при переходе от залов, отнесенных по общему звуковому впечатлению к высшей категории А, к залам третьей категории С. Усредненные значения параметров показаны в таблице 2:

Категория	1-КВСККр3	С80, дБ	СЗр3, дБ	СЗнч2, дБ	КНТ
А	0,66	-2,5	6,2	6,0	1,08
В	0,56	0,0	3,3	2,9	1,09
С	0,41	0,1	1,7	0,4	1,10

Способы, которыми обеспечиваются оптимальные значения объективных параметров, различаются в зависимости от назначения помещения: концертные залы, аппаратные звукозаписи, домашние помещения прослушивания и др.

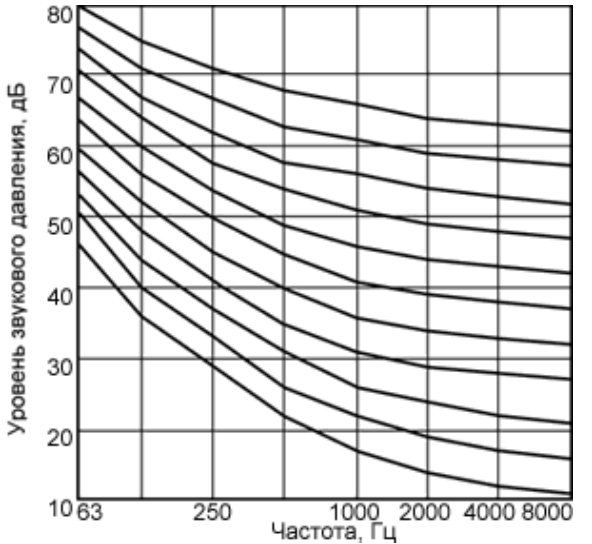
Полученные результаты имеют существенное значение, поскольку они позволили выявить ряд объективных параметров, обеспечивающих устойчивую корреляцию с субъективной оценкой качества звучания в различных залах. К их числу относятся, наряду со временем реверберации ВР, отношение ранней энергии к реверберирующей С80, интервал времени между прямым звуком и первыми отражениями t1, меры громкости СЗр3, СЗнч2, коэффициент низких тонов КНТ, коэффициент кросс-корреляции КВСККр3 и др.

Измерение этих коэффициентов в студиях, концертных залах, аудиториях и др. может оказать существенную помощь звукорежиссерам в выборе и при модернизации помещений для звукозаписи, а также при моделировании соответствующих условий с помощью компьютерных технологий.

Часть 14. Тембр, часть 1

Определения тембра

Как уже было отмечено в предыдущих статьях, одной из главных задач психоакустики является установление соответствия между объективными параметрами звука (интенсивность, длительность, периодичность, расположение в пространстве и др.) и его субъективно воспринимаемыми характеристикам (высота,



громкость, маскировка, тембр и др.). Как известно, связь между ними неоднозначна и нелинейна. Однако можно сказать, что субъективно ощущаемая высота тона связана в первую очередь с частотой (периодичностью), громкость - с интенсивностью и т. д. Высота тона позволяет классифицировать звуки по линейной шкале (выше-ниже), и служит в музыке основой мелодии, гармонии, интонации и пр. В свою очередь громкость определяет музыкальную динамику (ff...pp), баланс инструментов в ансамбле и является объемной характеристикой звука (больше-меньше).

Самым сложным субъективно ощущаемым параметром является тембр. С определением этого термина возникают сложности, сопоставимые с определением понятия "жизнь": все понимают, что это такое, однако над научным определением наука бьется уже несколько столетий. Аналогично с термином "тембр": всем ясно, о чем идет речь, когда говорят "красивый тембр голоса", "глухой тембр инструмента" и т. д., но... О тембре нельзя сказать "больше-меньше", "выше-ниже", для его описания используются десятки слов: сухой, звонкий, мягкий, резкий, яркий и т. д. (О терминах для описания тембра поговорим отдельно).

Уже более двухсот лет многие выдающиеся ученые пытаются дать научное определение этого параметра, которое, естественно, меняется с расширением наших представлений о механизмах работы слуховой системы. Определение тембра дается в трудах таких всемирно известных ученых, как Гельмгольц (1877), Флетчер (1938), Ликлайде (1951), Плом (1976), Наутсм (1989), Россин (1990), Ханде (1995).

Тембр (timbre-фр.) означает "качество тона", "окраску тона" (tone quality).

Американский стандарт ANSI-60-дает такое определение: "Тембр - атрибут слухового восприятия, который позволяет слушателю судить, что два звука, имеющие одинаковую высоту и громкость, различаются друг от друга".

В трудах Гельмгольца содержится следующее заключение: "разница в музыкальном качестве тона (тембре) зависит только от присутствия и силы парциальных тонов (обертонов), и не зависит от разности фаз, с которой эти парциальные тоны вступают в композицию". Это определение почти на сто лет определило направление исследований в области восприятия тембров, и претерпело существенные изменения и уточнения только в последние десятилетия. В трудах Гельмгольца был сделан еще целый ряд тонких наблюдений, которые подтверждаются современными результатами. В частности, им было установлено, что восприятие тембра зависит и от того, с какой скоростью парциальные тоны вступают в начале звука и умирают в его конце, а также, что наличие некоторых шумов и нерегулярностей помогает в распознавании тембров отдельных инструментов.

В 1938 г. Флетчер заметил, что тембр зависит от обертоновой структуры звука, но также изменяется при изменении громкости и высоты тона, хотя обертоновая структура может при этом сохраняться. В 1951 г. известный специалист Ликлайдер добавил, что тембр является многомерным объектом восприятия - он зависит от общей обертоновой структуры звука, которая также может меняться с изменением громкости и высоты тона.

В 1973 г. к определению тембра, данному в вышеприведенном стандарте ANSI, было сделано следующее добавление: "тембр зависит от спектра сигнала, но он также зависит от формы волны, звукового давления, расположения частот в спектре и временных характеристик звука".

Только к 1976 г. в работах Пломпа было доказано, что ухо не страдает "фазовой глухотой", и восприятие тембра зависит как от амплитудного спектра (в первую очередь, от формы спектральной огибающей), так и от фазового спектра. В 1990 году Россинг добавил, что тембр зависит от временной огибающей звука и его длительности. В работах 1993-1995 гг. отмечено, что тембр является субъективным атрибутом того или иного источника (например, голоса, музыкального инструмента), то есть он позволяет выделить этот источник из различных звуковых потоков в различных условиях. Тембр обладает достаточной инвариантностью (стабильностью), что позволяет сохранить его в памяти, а также служит для сравнения ранее записанной и вновь поступившей в слуховую систему информации об источнике звука. Это предполагает определенный процесс обучения - если человек никогда не слышал звучание инструмента данного тембра, то он его и не узнает.

За последние годы исследованию восприятия тембра и установлению его зависимостей от физических характеристик звука было посвящено сотни статей и книг. Большие научные коллективы ведущих университетов и институтов (Стенфордский университет, Гарвардский университет, университет Беркли, ИРКАМ и др.) ведут интенсивные исследования, поскольку результаты, полученные в этом направлении, являются принципиально важными для решения общей проблемы "расшифровки слухового образа", что необходимо в целях дальнейшего развития аудиотехники, систем мультимедиа и других направлений. Рассказать об огромном количестве полученных за последние годы результатов по изучению восприятия тембра в коротких статьях невозможно, поэтому постараюсь остановиться только на некоторых из них, поскольку они чрезвычайно важны для работы звукорежиссера со звуком.

Тембр и акустические характеристики звука

Современные компьютерные технологии позволяют выполнить детальный анализ временной структуры любого музыкального сигнала - это может сделать практически любой музыкальный редактор, например, Sound Forge, Wave Lab, SpectroLab и др. Примеры временной структуры (осциллограмм) звуков одной высоты

(нота "до" первой октавы), создаваемых различными инструментами (орган, скрипка), показаны на рисунке 1. Как видно из представленных волновых форм (т. е. зависимости изменения звукового давления от времени), в каждом из этих звуков можно выделить три фазы: атаку звука (процесс установления), стационарную часть, процесс спада. В различных инструментах, в зависимости от используемых в них способов звукообразования, временные интервалы этих фаз разные - это видно на рисунке 1.

У ударных и щипковых инструментов, например гитары, короткий временной отрезок стационарной фазы и атаки и длинный по времени - фазы затухания. В звуке органной трубы можно видеть достаточно длинный отрезок стационарной фазы и короткий период затухания и т. д. Если представить отрезок стационарной части звучания более растянутым во времени (Рис. 2), то можно отчетливо видеть периодическую структуру звука. Как уже было сказано в предыдущих статьях, эта периодичность является принципиально важной для определения музыкальной высоты тона, поскольку слуховая система только для периодических сигналов может определить высоту, а непериодические сигналы воспринимаются ею как шумовые.

Как утверждает классическая теория, развиваемая, начиная с Гельмгольца почти все последующие сто лет, восприятие тембра зависит от спектральной структуры звука, то есть от состава обертонов и соотношения их амплитуд. Позволю себе напомнить, что обертоны - это все составляющие спектра выше фундаментальной частоты, а обертоны, частоты которых находятся в целочисленных соотношениях с основным тоном, называются гармониками.

Как известно, для того, чтобы получить амплитудный и фазовый спектр, необходимо выполнить преобразование Фурье от временной функции (t), т. е. зависимости звукового давления p от времени t .

С помощью преобразования Фурье любой временной сигнал можно представить в виде суммы (или интеграла) составляющих его простых гармонических (синусоидальных) сигналов, а амплитуды и фазы этих составляющих образуют соответственно амплитудный и фазовый спектры.

С помощью созданных за последние десятилетия цифровых алгоритмов быстрого преобразования Фурье (БПФ или FFT), выполнить операцию по определению спектров можно также практически в любой программе обработки звука. Например, программа SpectroLab вообще является цифровым анализатором, позволяющим построить амплитудный и фазовый спектр музыкального сигнала в различной форме. Формы представления спектра могут быть различными, хотя представляют они одни и те же результаты расчетов. На рисунке 3 представлены в виде АЧХ амплитудные спектры различных музыкальных инструментов (осциллограммы которых были показаны на рисунке 2). АЧХ представляет здесь зависимость амплитуд обертонов в виде уровня звукового давления в дБ, от частот.

Иногда спектр представляют в виде дискретного набора обертонов с разными амплитудами. Спектры могут быть представлены в виде спектрограмм, где по вертикальной оси отложена частота, по горизонтальной - время, а амплитуда представлена интенсивностью цвета (Рис. 4). Кроме того, существует форма представления в виде трехмерного (кумулятивного) спектра, о котором будет сказано далее.

Для построения указанных на рисунке 3 спектров, в стационарной части осциллограммы выделяется некоторый временной отрезок, и проводится расчет усредненного спектра по данному отрезку. Чем больше этот отрезок, тем точнее получается разрешающая способность по частоте, но при этом могут теряться (сглаживаться) отдельные детали временной структуры сигнала. Такие стационарные спектры обладают индивидуальными чертами, характерными для каждого музыкального инструмента, и зависят от механизма звукообразования в нем.

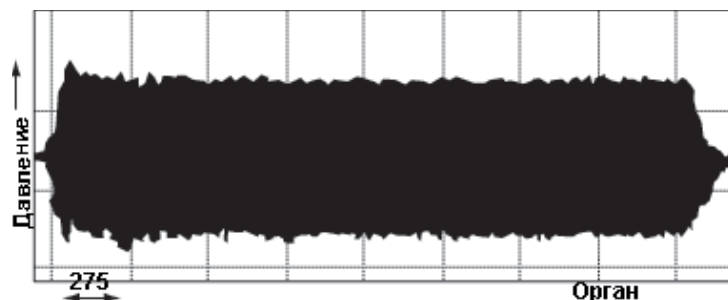


Рис. 1 Осциллограммы (волновая) форма звуков

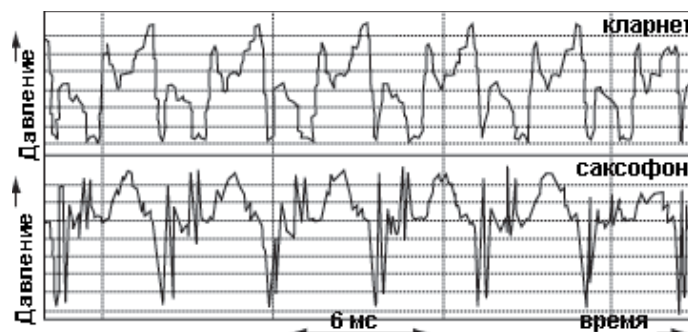
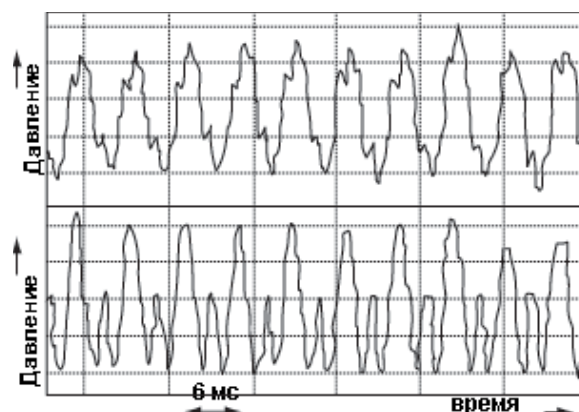


Рис. 2 Периодическая структура звуков.
а-орган; б-кларнет, саксофон

Например, флейта использует в качестве резонатора открытую с двух концов трубу, и поэтому содержит в спектре все четные и нечетные гармоники. При этом уровень (амплитуда) гармоник быстро уменьшается с частотой. У кларнета используется в качестве резонатора труба, закрытая с одного конца, поэтому в спектре, в основном, содержатся нечетные гармоники. У трубы в спектре много высокочастотных гармоник. Соответственно, тембры звучания у всех этих инструментов совершенно разные: у флейты - мягкий, нежный, у кларнета - матовый, глуховатый, у трубы - яркий, резкий.

Исследованию влияния спектрального состава обертонов на тембр посвящены сотни работ, поскольку эта проблема чрезвычайно важна как для проектирования музыкальных инструментов и высококачественной акустической аппаратуры, особенно в связи с развитием аппаратуры Hi-Fi и High-End, так и для слуховой оценки фонограмм и др. задач, встающих перед звукорежиссером. Накопленный огромный слуховой опыт наших замечательных звукорежиссеров - П.К. Кондрашина, В.Г. Динова, Е.В. Никульского, С.Г. Шугаля и др. - мог бы представить бесценные сведения по этой проблеме (особенно если бы они написали о нем в своих книгах, чего хотелось бы им пожелать).

Поскольку этих сведений чрезвычайно много и они часто противоречивы, приведем только некоторые из них.

Анализ общей структуры спектров различных инструментов, показанных на рисунке 5, позволяет сделать следующие выводы:

- при отсутствии или недостатке обертонов, особенно в нижнем регистре, тембр звука становится скучным, пустым - примером может служить синусоидальный сигнал от генератора;
- присутствие в спектре первых пяти-семи гармоник с достаточно большой амплитудой придает тембру полноту и сочность;
- ослабление первых гармоник и усиление высших гармоник (от шестой-седьмой и выше) придает тембру резкость, скрипучесть;

Анализ огибающей амплитудного спектра для различных музыкальных инструментов позволил установить (Кузнецов "Акустика музыкальных инструментов"):

- плавный подъем огибающей (увеличение амплитуд определенной группы обертонов) в области 200...700 Гц позволяет получить оттенки сочности, глубины;
- подъем в области 2,5...3 кГц придает тембру полетность, звонкость;
- подъем в области 3...4,5 кГц придает тембру резкость, пронзительность и др.

Одна из многочисленных попыток классифицировать тембровые качества в зависимости от спектрального состава звука приведена в вышеуказанной книге (Рис. 6).

Многочисленные эксперименты с оценкой качества звучания (а, следовательно, тембра) акустических систем позволили установить влияние различных пиков-провалов АЧХ на заметность изменения тембра. В частности, показано, что заметность зависит от амплитуды, расположения по частотной шкале и добротности пиков-провалов на огибающей спектра (т. е. на АЧХ). В средней области частот пороги заметности пиков, т. е. отклонения от среднего уровня, составляют 2...3 дБ, причем заметность изменения тембра на пиках больше, чем на провалах. Узкие по ширине провалы (менее 1/3 октавы) почти не заметны на слух - по-видимому, это объясняется тем, что именно такие узкие провалы вносит помещение в АЧХ различных звуковых источников, и слух к ним привык.

Существенное влияние оказывает группировка обертонов в формантные группы, особенно в области максимальной чувствительности слуха. Поскольку именно расположение формантных областей служит главным критерием различимости звуков речи, наличие формантных частотных

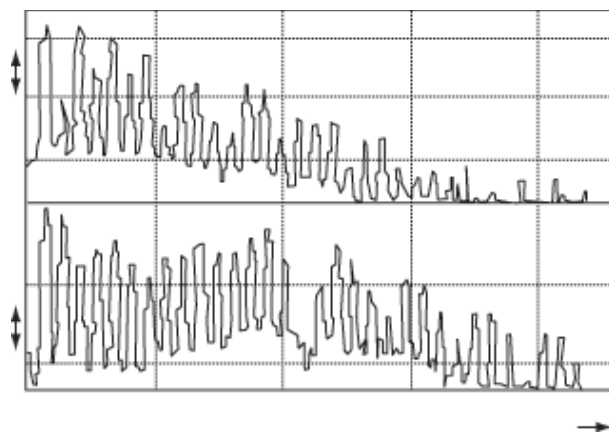
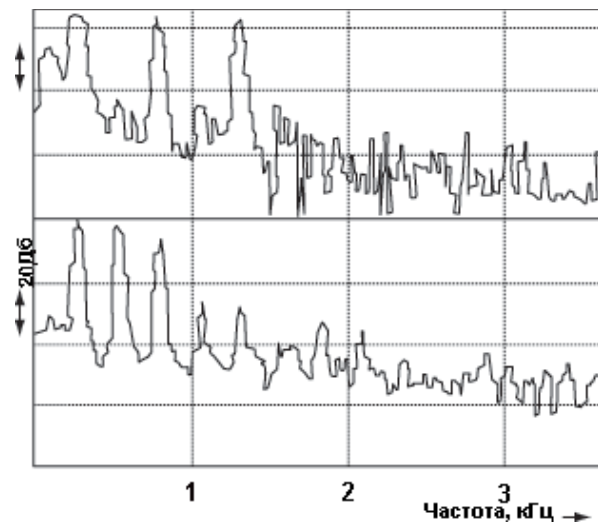


Рис. 3 Амплитудные спектры а-органа, б - кларнета, саксофона

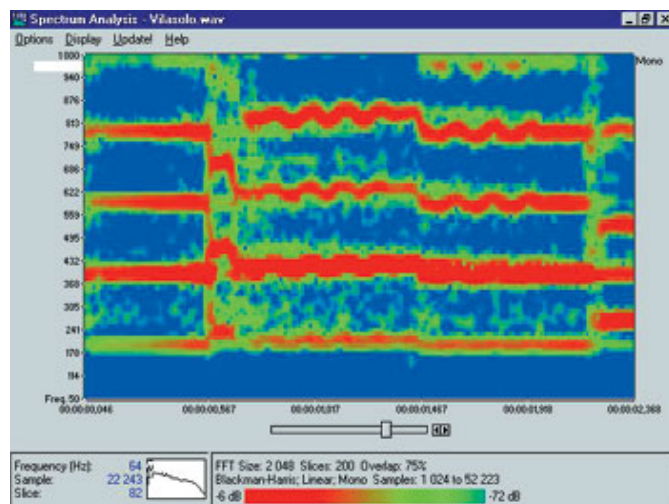


Рис. 4 Спектрограмма звука скрипки

диапазонов (т. е. подчеркнутых обертонов) значительно влияет на восприятие тембра музыкальных инструментов и певческого голоса: например, формантная группа в области 2...3 кГц придает полноту, звонкость певческому голосу и звукам скрипки. Эта третья форманта особенно выражена в спектрах скрипок Страдивари.

Таким образом, безусловно справедливо утверждение классической теории, что воспринимаемый тембр звука зависит от его спектрального состава, то есть расположения обертонов на частотной шкале и соотношения их амплитуд. Это подтверждается многочисленной практикой работы со звуком в разных областях. Современные музыкальные программы позволяют легко проверить это на простых примерах. Например, можно в Sound Forge синтезировать с помощью встроенного генератора варианты звуков с различным спектральным составом, и послушать, как изменяется тембр их звучания.

Из этого следуют еще два очень важных вывода:
- тембр звучания музыки и речи изменяется в зависимости от изменения громкости и от транспонирования по высоте.

При изменении громкости меняется восприятие тембра. Во-первых, при увеличении амплитуды колебаний вибраторов различных музыкальных инструментов (струн, мембран, дек и др.) в них начинают проявляться нелинейные эффекты, и это приводит к обогащению спектра дополнительными обертонами. На рисунке 7 показан спектр фортепиано при разной силе удара, где штрихом отмечена шумовая часть спектра.

Во-вторых, с увеличением уровня громкости изменяется чувствительность слуховой системы к восприятию низких и высоких частот (о кривых равной громкости было написано в предыдущих статьях). Поэтому при повышении громкости (до разумного предела 90...92 дБ) тембр становится полнее, богаче, чем при тихих звуках. При дальнейшем увеличении громкости начинают сказываться сильные искажения в источниках звука и слуховой системе, что приводит к ухудшению тембра.

Транспонирование мелодии по высоте также меняет воспринимаемый тембр. Во-первых, обедняется спектр, поскольку часть обертонов попадает в неслышимый диапазон выше 15...20 кГц; во-вторых, в области высоких частот пороги слуха значительно выше, и высокочастотные обертоны становятся не слышны. В звуках низкого регистра (например, в органе) обертоны усиливаются из-за повышения чувствительности слуха к средним частотам, поэтому звуки низкого регистра звучат сочнее, чем звуки среднего регистра, где такого усиления обертонов нет. Следует отметить, что поскольку кривые равной громкости, как и потери чувствительности слуха к высоким частотам, в значительной степени индивидуальны, то и изменение восприятия тембра при изменении громкости и высоты также очень различаются у разных людей.

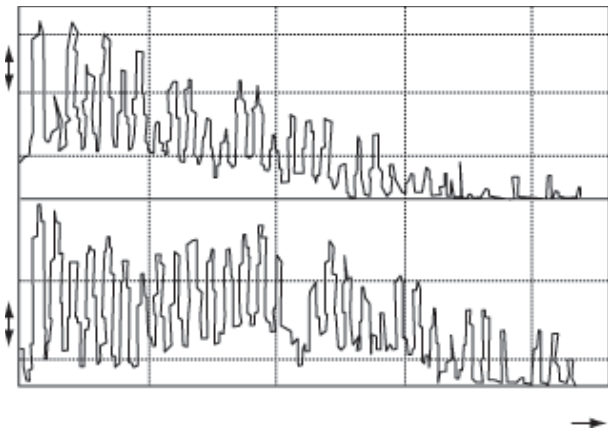


Рис. 5 Виды спектров различных инструментов

Однако, накопленные к настоящему времени экспериментальные данные позволили выявить определенную инвариантность (стабильность) тембра при целом ряде условий. Например, при транспонировании мелодии по частотной шкале оттенки тембра, конечно, меняются, но в целом тембр инструмента или голоса легко опознается: при прослушивании, например, саксофона или другого инструмента через транзисторный радиоприемник можно опознать его тембр, хотя спектр его был значительно искажен. При прослушивании одного и того же инструмента в разных точках зала его тембр так же меняется, но принципиальные свойства тембра, присущие данному инструменту, остаются.

Некоторые из этих противоречий удалось частично объяснить в рамках классической спектральной теории тембра. Например, было показано, что для сохранения основных признаков тембра при транспонировании (переносе по частотной шкале) принципиально важным является сохранение формы огибающей амплитудного спектра (т. е. его формантной структуры). Например, на рисунке 8 показано, что при переносе спектра на октаву в том случае, когда структура огибающей сохраняется (вариант "а"), вариации тембра менее значительны, чем при переносе спектра с сохранением соотношения амплитуд (вариант "б"). Этим объясняется то, что звуки речи (гласные, согласные) можно распознать независимо от того, с какой высотой (частотой фундаментального тона) они произнесены, если при этом сохраняется расположение их формантных



Рис. 6 Классификация тембров

областей относительно друг друга.

Таким образом, подводя итоги, полученные классической теорией тембра с учетом результатов последних лет, можно сказать, что тембр, безусловно, существенно зависит от усредненного спектрального состава звука: количества обертонов, их относительного расположения на частотной шкале, от соотношения их амплитуд, то есть формы спектральной огибающей (АЧХ), а точнее, от спектрального распределения энергии по частоте.

Однако, когда в 60-х годах начались первые опыты синтеза звуков музыкальных инструментов, попытки воссоздать звучание, в частности, трубы по известному составу ее усредненного спектра оказались неудачными - тембр был совершенно не похож на звук медных духовых инструментов. То же относится и к первым попыткам синтеза голоса. Именно в это период, опираясь на возможности, который предоставили компьютерные технологии, началось развитие другого направления - установление связи восприятия тембра с временной структурой сигнала.

Прежде, чем переходить к результатам, полученным в этом направлении, надо сказать следующее.

Первое. Довольно широко распространено мнение, что при работе со звуковыми сигналами достаточно получить информацию об их спектральном составе, поскольку перейти к их временной форме всегда можно с помощью преобразования Фурье, и наоборот. Однако, однозначная связь между временным и спектральным представлениями сигнала существует только в линейных системах, а слуховая система является принципиально нелинейной системой, как при больших, так и при малых уровнях сигнала. Поэтому обработка информации в слуховой системе происходит параллельно как в спектральной, так и во временной области (см. "Звукорежиссер" 6/1999 г.).

Разработчики высококачественной акустической аппаратуры сталкиваются с этой проблемой постоянно, когда искажения АЧХ акустической системы (то есть неравномерность спектральной огибающей) доведены почти до слуховых порогов (неравномерность 2 дБ, ширина полосы 20 Гц...20 кГц и т. д.), а эксперты или звукорежиссеры говорят: "скрипка звучит холодно" или "голос с металлом" и т.п. Таким образом, информации, полученной из спектральной области, для слуховой системы недостаточно, нужна информация о временной структуре. Неудивительно, что методы измерений и оценки акустической аппаратуры существенно изменились за последние годы - появилась новая цифровая метрология, позволяющая определить до 30 параметров, как во временной, так и в спектральной областях.

Следовательно, информацию о тембре музыкального и речевого сигнала слуховая система должна получать как из временной, так и из спектральной структуры сигнала.

Второе. Все полученные выше результаты в классической теории тембра (теории Гельмгольца) базируются на анализе стационарных спектров, полученных из стационарной части сигнала с определенным усреднением, однако принципиально важным является то обстоятельство, что в реальных музыкальных и речевых сигналах практически нет постоянных, стационарных частей. Живая музыка - это непрерывная динамика, постоянное изменение, и это связано с глубинными свойствами слуховой системы.

Исследования физиологии слуха позволили установить, что в слуховой системе, особенно в ее высших разделах, имеется множество так называемых нейронов "новизны" или "опознавания", т. е. нейронов, которые включаются и начинают проводить электрические разряды, только если есть изменения в сигнале (включение, выключение, изменение уровня громкости, высоты и т. д.). Если же сигнал стационарный, то эти нейроны не включаются, и контроль за сигналом осуществляет ограниченное количество нейронов. Это явление широко известно из повседневной жизни: если сигнал не меняется, то часто его просто перестают замечать.

Для музыкального исполнения всякие монотонность и постоянство являются губительными: у слушателя отключаются нейроны новизны и он перестает воспринимать информацию (эстетическую, эмоциональную, смысловую и др), поэтому в живом исполнении всегда есть динамика (музыканты и певцы широко используют различную модуляцию сигнала - вибрато, тремоло и пр.).

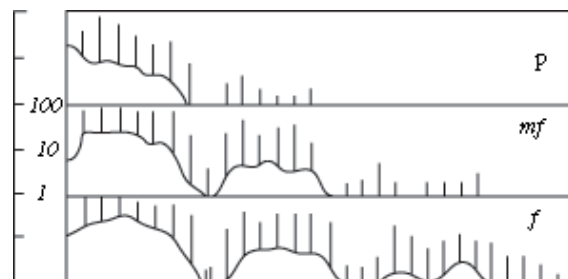


Рис. 7 Зависимость состава спектра от силы удара

Кроме того, каждый музыкальный инструмент, включая голос, обладает особой системой звукообразования, которая диктует свою временную структуру сигнала и его динамику изменения. Сравнение временной структуры звука (Рис. 1) показывает принципиальные различия: в частности, длительности всех трех частей - атаки, стационарной части и спада - у всех инструментов различаются по продолжительности и по форме. У ударных инструментов очень короткая стационарная часть, время атаки 0,5...3 мс и время спада 0,2...1 с; у смычковых время атаки 30...120 мс, время спада 0,15...0,5 с; у органа атака - 50...1000 мс и спад 0,2...2 с. Кроме того, принципиально отличается форма временной огибающей (Рис. 1).

Эксперименты показали, что, если удалить часть временной структуры, соответствующей атаке звука, или поменять местами атаку и спад (проиграть в обратном направлении), или атаку от одного инструмента заменить атакой от другого, то опознать тембр данного инструмента становится практически невозможным. Следовательно, для распознавания тембра не только стационарная часть (усредненный спектр которой служит основой классической теории тембра), но и период формирования временной структуры, как и период затухания (спада) являются жизненно важными элементами.

Действительно, при прослушивании в любом помещении первые отражения поступают на слуховую систему после того, как атака и начальная часть стационарной части уже была услышана. В то же время на спад звука от инструмента накладывается реверберационный процесс помещения, что значительно маскирует звук, и, естественно, приводит к модификации восприятия его тембра. Слух обладает определенной инерционностью, и короткие звуки воспринимаются как щелчки. Поэтому длительность звука должна быть больше 60 мс, чтобы можно было распознать высоту, и, соответственно, тембр. По-видимому, постоянные должны быть близки.

Тем не менее, времени между началом прихода прямого звука и моментами поступления первых отражений оказывается достаточно, чтобы распознать тембр звучания отдельного инструмента - очевидно, этим обстоятельством и определяется инвариантность (стабильность) распознавания тембров разных инструментов в разных условиях прослушивания. Современные компьютерные технологии позволяют достаточно детально проанализировать процессы установления звука у разных инструментов, и выделить самые существенные акустические признаки, наиболее важные для определения тембра. Детальный анализ этих признаков будет выполнен во второй части статьи.

Часть 14 Тембр. Часть 2

Как уже было отмечено в первой части этой статьи (["Звукорежиссер", 2/2001](#)), существенное влияние на восприятие тембра музыкального инструмента или голоса оказывает структура его стационарного (усредненного) спектра: состав обертонов, их расположение на частотной шкале, их частотные соотношения, распределения амплитуд и форма огибающей спектра, наличие и форма формантных областей и т.д., что полностью подтверждает положения классической теории тембра, изложенные еще в трудах Гельмгольца. Однако экспериментальные материалы, полученные за последние десятилетия, показали, что не менее существенную, а, может быть, и гораздо более существенную роль в распознавании тембра играет нестационарное изменение структуры звука и, соответственно, процесс развертывания во времени его спектра, в первую очередь, на начальном этапе атаки звука.

Процесс изменения спектра во времени особенно наглядно можно "увидеть" с помощью спектрограмм или трехмерных спектров (они могут быть построены с помощью большинства музыкальных редакторов Sound Forge, SpectroLab, Wave Lab и др.). Их анализ для звуков различных инструментов позволяет выявить характерные особенности процессов "развертывания" спектров. Например, на рисунке 1 показан трехмерный спектр звучания колокола, где по одной оси отложена частота в Гц, по другой время в секундах; по третьей амплитуда в дБ. На графике отчетливо видно, как происходит процесс нарастания, установления и спада во времени спектральной огибающей.

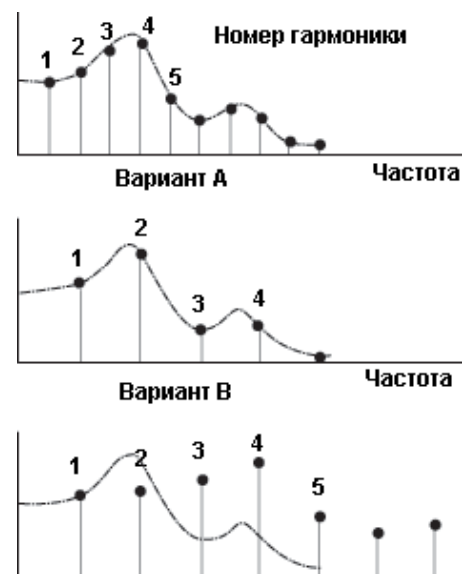


Рис. 8 Изменение огибающей спектра при сдвиге высоты

Рис.1. Трехмерный спектр звучания колокола

Процесс атаки у большинства музыкальных инструментов и голоса продолжается несколько десятков миллисекунд. За этот период времени (иногда с переходом на стационарную часть) слух воспринимает постепенное расширение спектра во времени, поскольку вступают все новые обертоны с различной скоростью и амплитудой, и распознает тембр данного инструмента. На этот процесс распознавания оказывают влияние также многочисленные другие признаки: начальный скрип смычки, начало ноты на медном инструменте, шум дыхания, начальное прохлопывание язычка, ударный глухой звук молоточка на пианино, небольшая негармоничность обертонов и т.д. все это создает живые акустические признаки идентификации инструмента (об их роли поговорим позднее).

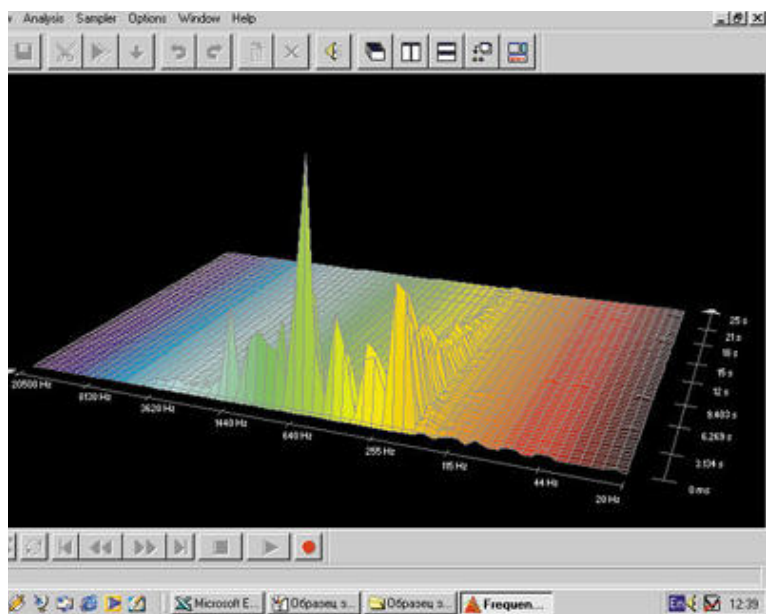


Рис.2. Атака звука для закрытой и открытой лабиальной (флейтовой) органной трубы

Как известно, процесс атаки особенно важен для распознавания тембра также еще и потому, что он является устойчивой характеристикой звучания данного инструмента, менее всего подверженной "окрашиванию" со стороны помещения, в котором данное произведение исполняется, поскольку первые отражения поступают к слушателю с определенным запаздыванием, после того, как фаза атаки звука уже завершена и поступила к слушателю неокрашенной в виде прямого звука. Если бы этого не было, то распознать тембр инструмента при исполнении в различных помещениях было бы практически невозможно. Эксперименты показали, что слушатели не могут распознать инструмент, если фаза атаки удалена или изменена. Если поменять фазы атаки и спада местами (проиграть, например, запись любого инструмента в обратном направлении), то тембр меняется до неузнаваемости.

Измерения показывают существенные различия в структуре атаки у разных инструментов. На рисунках 2...4 показаны примеры уровнеграмм (зависимостей уровня звукового давления от времени) и их спектрограмм (зависимостей частоты обертонов от времени, амплитуда при этом отмечается яркостью) для начального периода установления колебаний (атаки) тона С4 для различных инструментов: органнх труб, деревянных духовых и медных духовых.

Сравнение спектрограмм для двух типов органнх труб открытая флейтовая (principal 8', рис. 2) и закрытая флейтовая (Gedakt 8') показывает, что у открытой трубы первой в спектре начинает устанавливаться вторая гармоника (на октаву выше первой). Она опережает первую гармонику (основную частоту) почти на 30 мс, почти одновременно с основным тоном появляется третья гармоника, четвертая и выше гармоники малы по амплитуде. У закрытой трубы первой в спектре появляется пятая гармоника, на две октавы + большую терцию выше первой; затем появляется первая гармоника; и только затем третья, на дуодециму (октаву + квинту) выше первой. Такое различие в процессе атаки отражает физический механизм звукообразования в разных трубах, поскольку у открытых труб возбуждаются все гармоники, а у закрытых только нечетные, что определяет различный тембр их звучания.

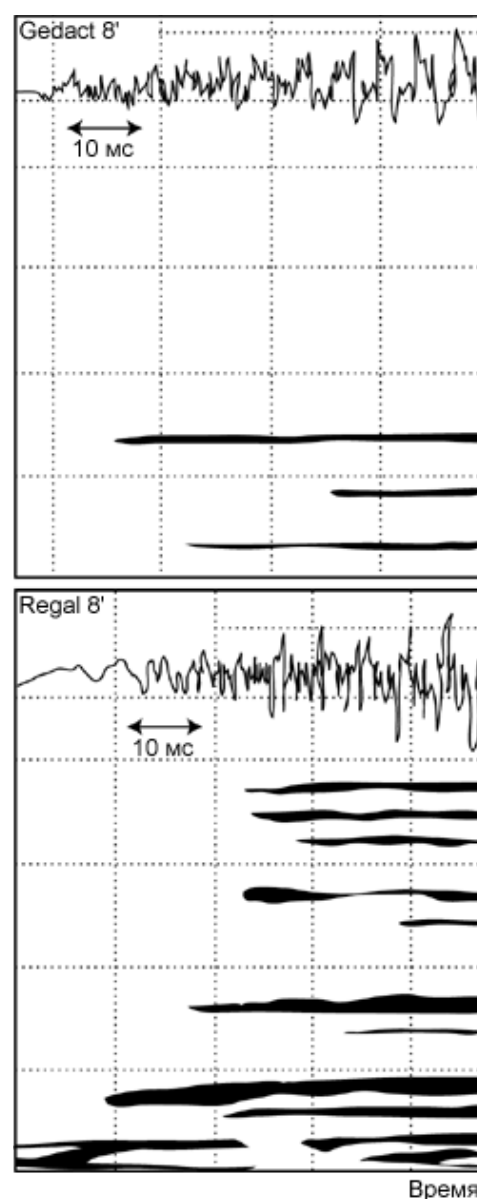


Рис.3. Атака звука для деревянных духовых инструментов: кларнет, гобой, флейта

Сравнение атаки тона С4 у различных деревянных инструментов показывает, что процесс установления колебаний у каждого инструмента имеет свой особый характер (рис. 3):

- у кларнета доминируют нечетные гармоники 1/3/5, причем третья гармоника появляется в спектре на 30 мс позже первой, затем постепенно "выстраиваются" более высокие гармоники;
- у гобоя установление колебаний начинается со второй и третьей гармоники, затем появляется четвертая и только через 8 мс начинает появляться первая гармоника;
- у флейты сначала появляется первая гармоника, затем только через 80 мс постепенно вступают все остальные.

На рисунке 4 показан процесс установления колебаний для группы медных инструментов: трубы, тромбона, валторны и тубы. Отчетливо видны различия:

- у трубы компактное появление группы высших гармоник, у тромбона первой появляется вторая гармоника, затем первая, и через 10 мс вторая и третья. У тубы и валторны видна концентрация энергии в первых трех гармониках, высшие гармоники практически отсутствуют.

Рис.4. Атака звука для медных духовых инструментов: труба, тромбон, валторна, туба

Анализ полученных результатов показывает, что процесс атаки звука существенно зависит от физической природы звукоизвлечения на данном инструменте:

- от использования амбушюров или тростей, которые, в свою очередь, делятся на одинарные или двойные;
- от различных форм труб (прямые узкомензурные или конусные широкомензурные) и т.д.

Это определяет количество гармоник, время их появления, скорость выстраивания их амплитуды, а соответственно и форму огибающей временной структуры звука. У некоторых инструментов, например, флейты (Рис. 5а), огибающая в период атаки имеет плавный экспоненциальный характер, а у некоторых, например, фагота (Рис.5б), отчетливо видны биения, что и является одной из причин существенных различий в их тембре.

Во время атаки высшие гармоники иногда опережают основной тон, поэтому могут происходить флуктуации высоты тона периодичность, а значит, и высота суммарного тона, выстраиваются постепенно. Иногда эти изменения периодичности носят квазислучайный характер. Все эти признаки помогают слуховой системе "опознать" тембр того или иного инструмента в начальный момент звучания.

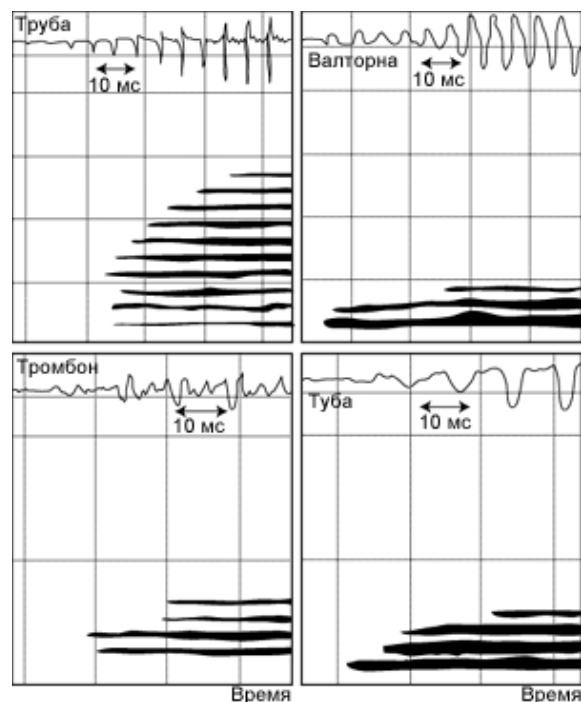
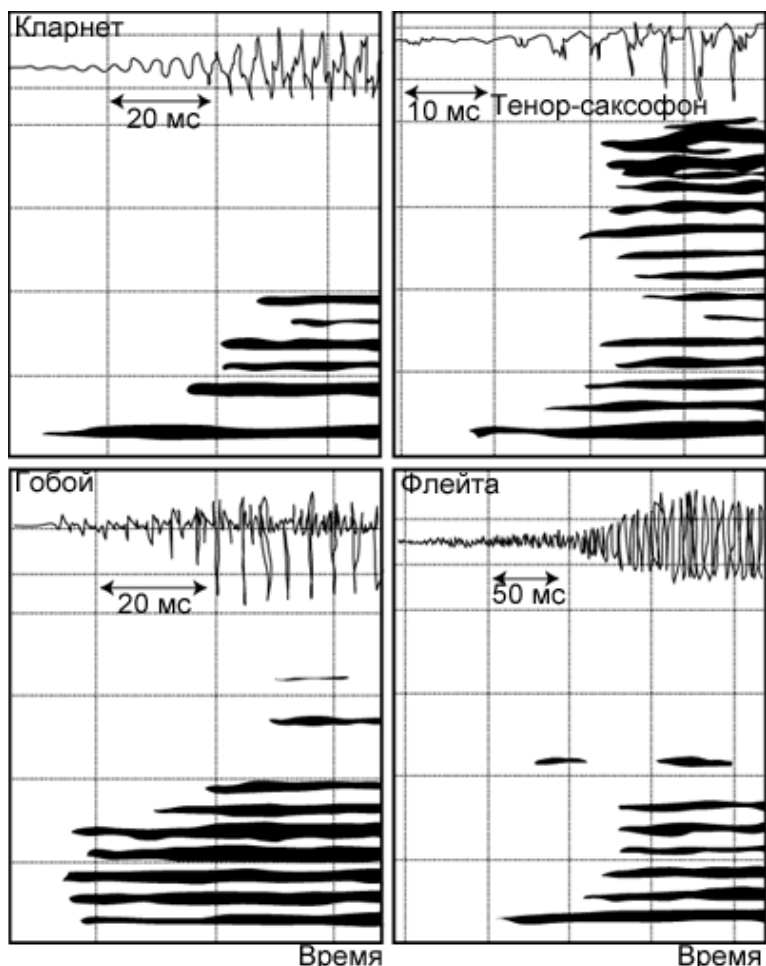
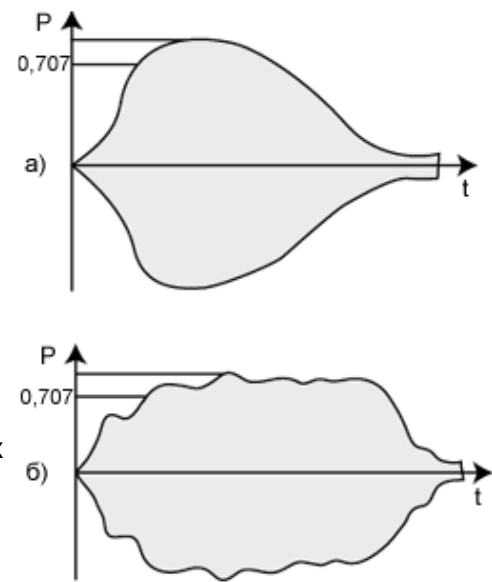


Рис.5. Временная огибающая атаки: а) флейты, б) фагота



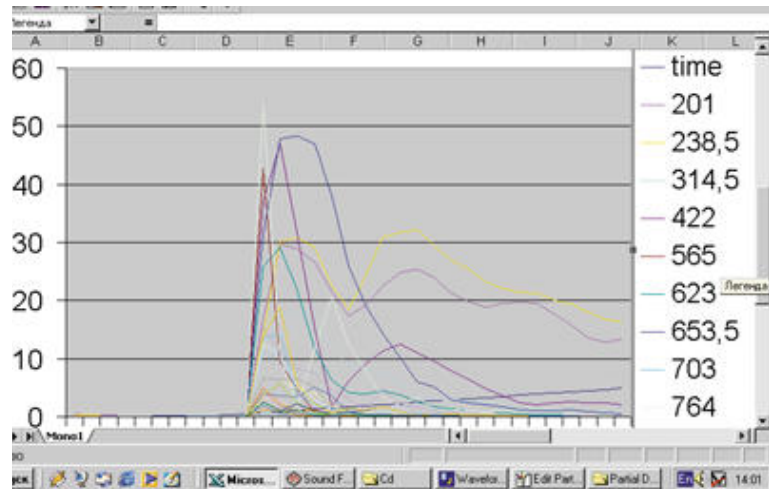
Для оценки тембра звучания важен не только момент его распознавания (т.е. способность отличить один инструмент от другого), но и возможность оценить изменение тембра в процессе исполнения. Здесь важнейшую роль играет динамика изменения спектральной огибающей во времени на всех этапах звучания: атаки, стационарной части, спада.

Характер поведения каждого обертона во времени также несет важнейшую информацию о тембре. Например, в звучании колоколов особенно четко видна динамика изменения, как по составу спектра, так и по характеру изменения во времени амплитуд его отдельных обертонов: если в первый момент после удара в спектре отчетливо видно несколько десятков спектральных составляющих, что создает шумовой характер тембра, то через несколько секунд в спектре остаются несколько основных обертонов (основной тон, октава, дуодецима и минорная терция через две октавы), остальные затухают, и это создает особый тонально окрашенный тембр звучания.

Рис.6. Изменения во времени амплитуд обертонов для звука колокола

Пример изменения амплитуд основных обертонов во времени для колокола показан на рисунке 6. Видно, что для него характерна короткая атака и длинный период затухания, при этом скорость вступления и спада обертонов различных порядков и характер изменения их амплитуд во времени существенно отличаются. Поведение различных обертонов во времени зависит от типа инструмента: в звучании рояля, органа, гитары и др. процесс изменения амплитуд обертонов имеет совершенно разный характер.

Опыт показывает, что аддитивный компьютерный синтез звуков, учитывающий специфику развертывания отдельных обертонов во времени, позволяет получить значительно более "жизненное" звучание.



Вопрос о том, динамика изменения каких именно обертонов несет информацию о тембре, связан с существованием критических полос слуха. Как уже было сказано в первой статье цикла ("Звукорежиссер", 6/1999), посвященной определению высоты тона, периферическая слуховая система выполняет спектральный анализ поступившего звука. Базиллярная мембрана в улитке действует как линейка полосовых фильтров, ширина полосы которых зависит от частоты: выше 500 Гц она равна примерно 1/3 октавы, ниже 500 Гц она составляет примерно 100 Гц. Ширина полосы этих слуховых фильтров называется "критической полосой слуха" (существует специальная единица измерения 1 барк, равная ширине критической полосы во всем диапазоне слышимых частот).

Внутри критической полосы слух производит интегрирование поступившей звуковой информации, что играет также важную роль в процессах слуховой маскировки. Если проанализировать сигналы на выходе слуховых фильтров, то можно видеть, что первые пять-семь гармоник в спектре звучания любого инструмента попадают обычно каждая в свою критическую полосу, поскольку они достаточно далеко отстоят друг от друга в таких случаях говорят, что гармоники "развертываются" слуховой системой. Разряды нейронов на выходе таких фильтров синхронизированы с периодом каждой гармоники.

Рис.7. Вид сигналов на выходе слуховых фильтров для "развернутых" и "неразвернутых" гармоник

Гармоники выше седьмой обычно находятся достаточно близко друг к другу по частотной шкале, и не "развертываются" слуховой системой внутри одной критической полосы попадает несколько гармоник, а на выходе слуховых фильтров получается сложный сигнал. Разряды нейронов в этом случае синхронизированы с частотой огибающей, т.е. основного тона (рис. 7). Соответственно, механизм обработки информации слуховой системой для развернутых и неразвернутых гармоник несколько отличается в первом случае используется информация "по времени", во втором "по месту".

Существенную роль при распознавании высоты тона, как было показано в предыдущих статьях, играют первые пятнадцать-восемнадцать гармоник. Эксперименты с помощью компьютерного аддитивного синтеза звуков показывают, что поведение именно этих гармоник оказывает также наиболее существенное влияние на изменение тембра.

Поэтому в ряде исследований предлагалось размерность тембра считать равной пятнадцати-восемнадцати, и оценивать его изменение по этому количеству шкал это одно из принципиальных отличий тембра от таких характеристик слухового восприятия, как высота или громкость, которые могут быть шкалированы по двум-трем параметрам (например, громкость), зависящих в основном от интенсивности, частоты и длительности сигнала.

Достаточно хорошо известно, что если в спектре сигнала присутствует достаточно много гармоник с номерами от 7-ой до 15...18-ой, с достаточно большими амплитудами, например, у трубы, скрипки, язычковых труб органа и т.п., то тембр воспринимается как яркий, звонкий, резкий и т. д. Если в спектре присутствуют в основном низшие гармоники, например, у тубы, валторны, тромбона, то тембр характеризуется как темный, глухой и т.д. (см. рис. 2...4). Кларнет, у которого в спектре доминируют нечетные гармоники, обладает несколько "носовым" тембром и т.д.

В соответствии с современными взглядами, важнейшую роль для восприятия тембра имеет изменение динамики распределения максимума энергии между обертонами спектра.

Рис.8. Определение центроида

Для оценки этого параметра введено понятие "центроид спектра", который определяется как средняя точка распределения спектральной энергии звука, его иногда определяют как "балансную точку" спектра. Способ определения его состоит в том, что рассчитывается значение некоторой средней частоты:

$F = \sum A_i f_i / \sum f_i$, где A_i амплитуда составляющих спектра, f_i их частота. Для примера, показанного на рисунке 8, это значение центроида составляет 200 Гц.

$F = (8 \times 100 + 6 \times 200 + 4 \times 300 + 2 \times 400) / (8 + 6 + 4 + 2) = 200.$

Смещение центроида в сторону высоких частот ощущается как повышение яркости тембра.

Существенное влияние распределения спектральной энергии по частотному диапазону и ее изменения во времени на восприятие тембра связано, вероятно, с опытом распознавания звуков речи по формантным признакам, которые и несут информацию о концентрации энергии в различных областях спектра (неизвестно, правда, что было первичным).

Эта способность слуха имеет существенное значение при оценке тембров музыкальных инструментов, поскольку наличие формантных областей характерно для большинства музыкальных инструментов, например, у скрипок в областях 800...1000 Гц и 2800...4000 Гц, у кларнетов 1400...2000 Гц и т.д. Соответственно, их положение и динамика изменения во времени влияют на восприятие индивидуальных особенностей тембра.

Известно, какое значительное влияние на восприятие тембра певческого голоса оказывает наличие высокой певческой форманты (в области 2100...2500 Гц у басов, 2500...2800 Гц у теноров, 3000...3500 Гц у сопрано). В этой области у оперных певцов сосредоточивается до 30% акустической энергии, что обеспечивает звонкость и полетность голоса. Удаление с помощью фильтров певческой форманты из записей различных голосов (эти опыты были выполнены в исследованиях проф. В.П. Морозова) показывает, что тембр голоса становится тусклым, глуховатым и вялым.

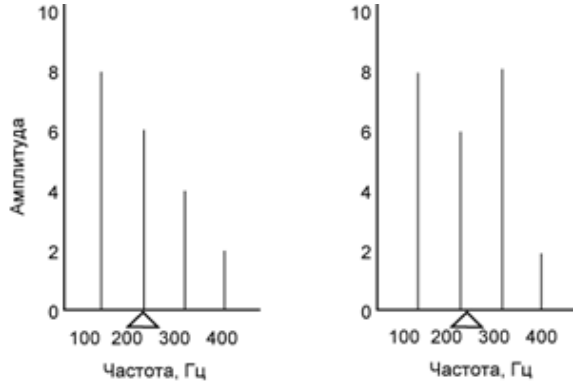
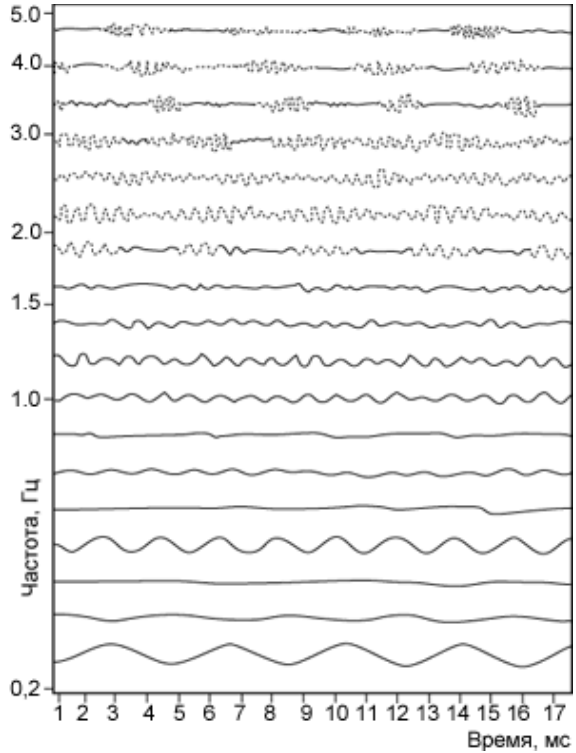
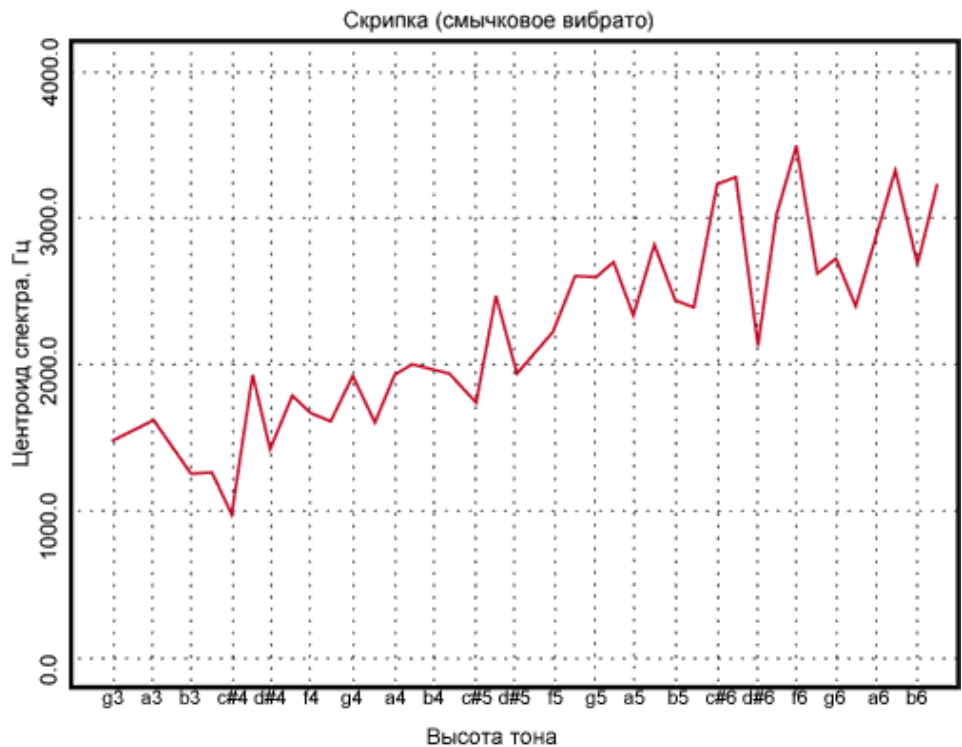


Рис.9. Изменение положения центроида для звуков скрипки

Изменение тембра при изменении громкости исполнения и транспонировании по высоте также сопровождается сдвигом центроида за счет изменения количества обертонов. Пример изменения положения центроида для звуков скрипки разной высоты показан на рисунке 9 (по оси абсцисс отложена частота расположения центроида в спектре). Исследования показали, что у многих музыкальных инструментов имеется почти монотонная связь между увеличением интенсивности (громкости) и сдвигом центроида в высокочастотную область, за счет чего тембр становится ярче.

По-видимому, при синтезе звуков и создании различных компьютерных композиций следует учитывать динамическую связь между интенсивностью и положением центроида в спектре для того, чтобы получать более естественный тембр.



Наконец, различие в восприятии тембров реальных звуков и звуков с "виртуальной высотой", т.е. звуков, высоту которых мозг "достраивает" по нескольким целочисленным обертонам спектра (это характерно, например, для звуков колоколов), можно объяснить с позиций положения центроида спектра. Поскольку у этих звуков значение частоты основного тона, т.е. высоты, может быть одинаковым, а положение центроида разное из-за разного состава обертонов, то, соответственно, тембр будет восприниматься по-разному.

Интересно отметить, что еще более десяти лет назад для измерения акустической аппаратуры был предложен новый параметр, а именно трехмерный спектр распределения энергии по частоте и по времени, так называемое распределение Вигнера, которое достаточно активно используется различными фирмами для оценки аппаратуры, поскольку, как показывает опыт, позволяет установить наилучшее соответствие с ее качеством звучания. Учитывая изложенное выше свойство слуховой системы использовать динамику изменения энергетических признаков звукового сигнала для определения тембра, можно предположить, что этот параметр распределение Вигнера может быть полезен и для оценки музыкальных инструментов.

Оценка тембров различных инструментов всегда носит субъективный характер, но если при оценке высоты и громкости можно на основе субъективных оценок расположить звуки по определенной шкале (и даже ввести специальные единицы измерения "сон" для громкости и "мел" для высоты), то оценка тембра значительно более трудная задача. Обычно для субъективной оценки тембра слушателям предъявляются пары звуков, одинаковых по высоте и громкости, и их просят расположить эти звуки по разным шкалам между различными противоположными описательными признаками: "яркий"/"темный", "звонкий"/"глухой" и т.д. (О выборе различных терминов для описания тембров и о рекомендациях международных стандартов по этому вопросу мы обязательно поговорим в дальнейшем).

Существенное влияние на определение таких параметров звука, как высота, тембр и др., оказывает поведение во времени первых пяти-семи гармоник, а также ряда "неразвернутых" гармоник до 15...17-ой. Однако, как известно из общих законов психологии, кратковременная память человека может одновременно оперировать не более чем семью-восемью символами. Поэтому очевидно, что и при распознавании и оценке тембра используется не более семи-восемью существенных признаков.

Попытки установить эти признаки путем систематизации и усреднения результатов экспериментов, найти обобщенные шкалы, по которым можно было бы идентифицировать тембры звуков различных инструментов, связать эти шкалы с различными временно-спектральными характеристиками звука, предпринимаются уже давно.

Одной из самых известных является работа Грея (1977 г.), где было проведено статистическое сравнение оценок по различным признакам тембров звуков различных инструментов струнных, деревянных, перкуссионных и др. Звуки были синтезированы на компьютере, что позволяло менять в требуемых направлениях их временные и спектральные характеристики. Классификация тембральных признаков была выполнена в трехмерном (ортогональном) пространстве, где в качестве шкал, по которым производилась сравнительная оценка степени подобия тембральных признаков (в пределах от 1 до 30), были выбраны следующие:

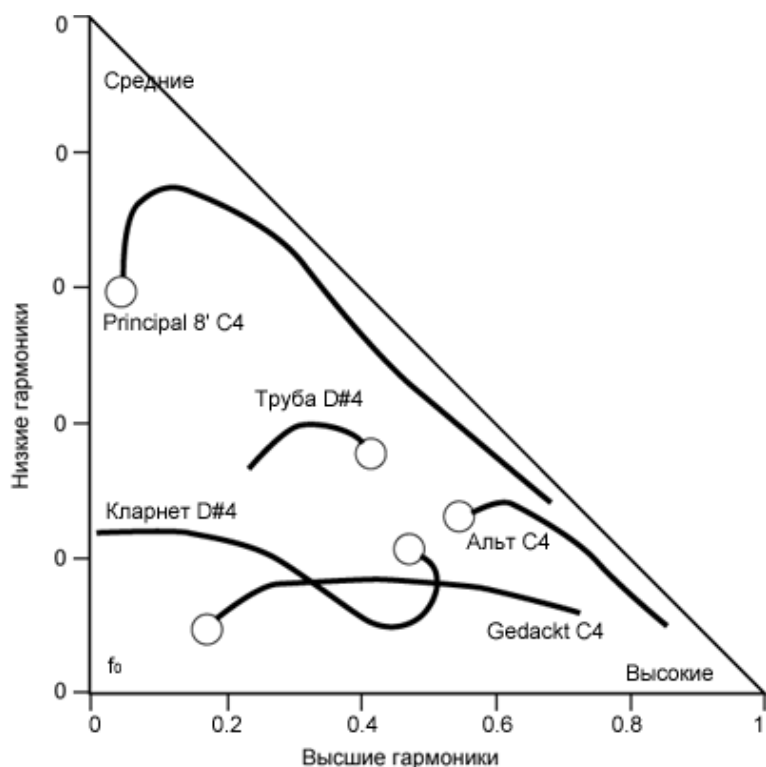
- первая шкала - значение центроида амплитудного спектра (по шкале отложено смещение центроида, т.е.

- максимума спектральной энергии от низких к высоким гармоникам);
- вторая - синхронность спектральных флуктуаций, т.е. степень синхронности вступления и развития отдельных обертонов спектра;
- третья - степень наличия низкоамплитудной негармонической высокочастотной энергии шума в период атаки.

Обработка полученных результатов с помощью специального пакета программ для кластерного анализа позволила выявить возможность достаточно четкой классификации инструментов по тембрам внутри предложенного трехмерного пространства .

Рис.10. Двухмерная диаграмма для классификации тембров инструментов

Попытка визуализировать тембральное различие звуков музыкальных инструментов в соответствии с динамикой изменения их спектра в период атаки была предпринята в работе Полларда (1982 г.), результаты показаны на рисунке 10. По оси Y отложена величина, пропорциональная общей энергии низших ("развернутых") гармоник, по оси X энергии высших "неразвернутых" гармоник. Если рассмотреть, как меняется энергия, приходящаяся на долю гармоник в период атаки звука (который продолжается от 10 до 160 мс для кларнета, от 10 до 65 мс для скрипки, от 10...100 мс для трубы), то можно достаточно четко разделить звуки различных инструментов по этим признакам. Процесс изменения соотношения гармоник в разные периоды атаки показан черной линией, переход к стационарному состоянию белым кружком. У кларнета в начальный период превалирует спектральная энергия в низших гармониках (0,25), затем увеличивается доля энергии высших гармоник (до 0,55), затем, при переходе к стационарному состоянию (отмечено кружком), устанавливается определенное соотношение в распределении энергии между низшими и высшими гармониками (0,2/0,45).



У скрипки в начальный период превалируют (энергетически) высокие гармоники, затем, по мере перехода к стационарному состоянию, уменьшается доля высоких гармоник, возрастает энергия в низших гармониках. Особенно наглядно видно перемещение энергетического максимума от высших гармоник в начальный момент атаки к низшим гармоникам по мере перехода к стационарному состоянию у органной трубы (principal 8'). Как следует из данных, приведенных на рисунке 10, все инструменты достаточно четко разделяются по этим спектральным признакам, что еще раз подчеркивает важность динамики изменения спектра в период атаки для идентификации тембра различных инструментов.

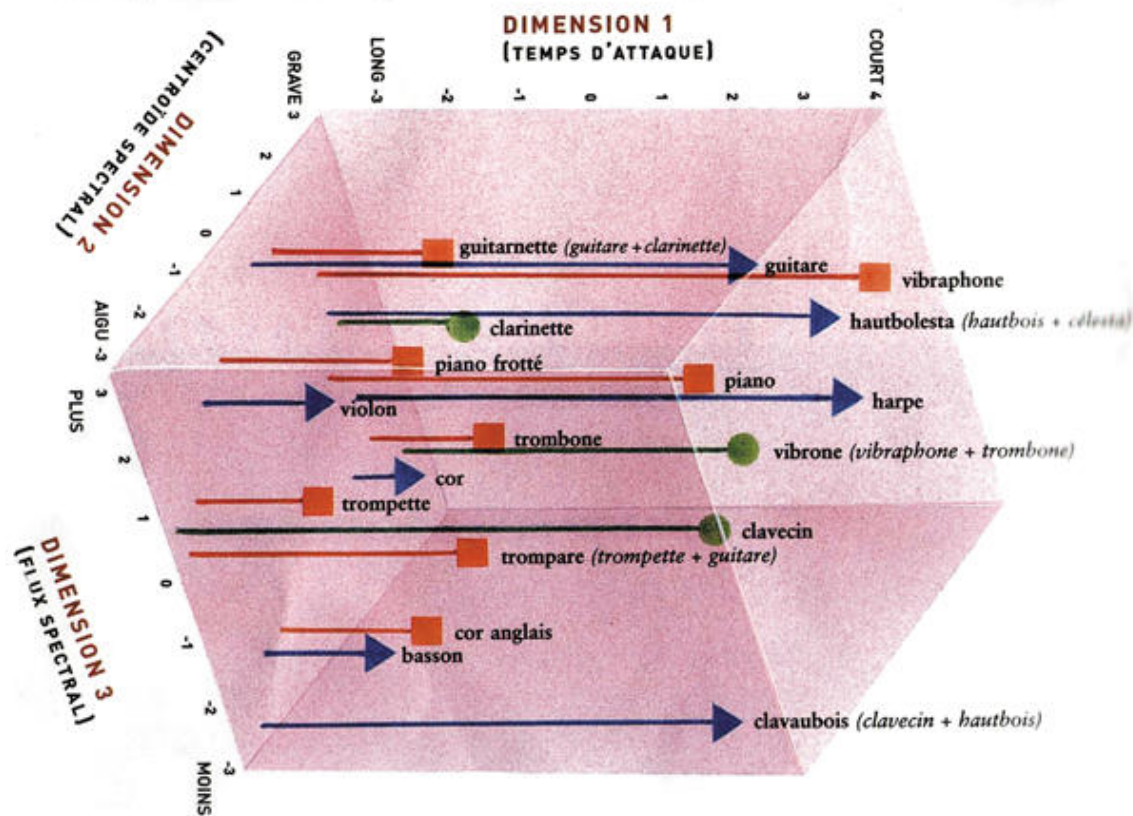


Рис.11. Трехмерное пространство тембров

Поиски методов многомерного шкалирования тембров и установление их связей с спектрально-временными характеристиками звуков активно продолжают. Эти результаты чрезвычайно важны для развития технологий компьютерного синтеза звуков, для создания различных электронных музыкальных композиций, для коррекции и обработки звука в звукорежиссерской практике и т.д.

Интересно отметить, что еще в начале века великий композитор XX века Арнольд Шёнберг высказал идею, что "...если рассматривать высоту тона, как одну из размерностей тембра, а современную музыку построенной на вариации этой размерности, то почему бы не попробовать использовать другие размерности тембра для создания композиций". Эта идея реализуется в настоящее время в творчестве композиторов, создающих спектральную (электроакустическую) музыку. Именно поэтому интерес к проблемам восприятия тембра и его связям с объективными характеристиками звука настолько высок.

Таким образом, полученные результаты показывают, что, если в первый период изучения восприятия тембра (на основе классической теории Гельмгольца) была установлена четкая связь изменения тембра с изменением спектрального состава стационарной части звучания (составом обертонов, соотношением их частот и амплитуд и др.), то второй период этих исследований (с начала 60-х годов) позволил установить принципиальную важность спектрально-временных характеристик.

Это изменение структуры временной огибающей на всех этапах развития звука: атаки (что особенно важно для распознавания тембров различных источников), стационарной части и спада. Это и динамическое изменение во времени спектральной огибающей, в т.ч. смещение центроида спектра, т.е. смещение максимума спектральной энергии во времени, а также развитие во времени амплитуд спектральных составляющих, особенно первых пяти-семи "неразвернутых" гармоник спектра.

В настоящее время начался третий период изучения проблемы тембра центр исследований переместился в сторону изучения влияния фазового спектра, а также к использованию психофизических критериев в распознавании тембров, лежащих в основе общего механизма распознавания звукового образа (группировка в потоки, оценка синхронности и др.). О полученных в этом направлении результатах в следующих статьях.

Часть 14. Тембр, часть 3

Тембр и фазовый спектр

Все изложенные в предыдущих двух статьях результаты по установлению связи воспринимаемого тембра с акустическими характеристиками сигнала относились к амплитудному спектру, точнее, к временному изменению спектральной огибающей (в первую очередь смещению энергетического центра амплитудного спектра-центроида) и разворачиванию во времени отдельных обертонов.

В этом направлении было проделано наибольшее количество работ и получено много интересных результатов. Как уже было отмечено, на протяжении почти ста лет в психоакустике превалировало мнение Гельмгольца о том, что наша слуховая система не чувствительна к изменениям фазовых соотношений между отдельными обертонами. Однако постепенно были накоплены экспериментальные данные о том, что слуховой аппарат чувствителен к изменениям фаз между различными компонентами сигнала (работы Шредера, Хартмана и др.). В частности, было установлено, что слуховой порог к фазовому сдвигу в двух- и трехкомпонентных сигналах в области низких и средних частот составляет 10...15 градусов.

В 80-х годах это привело к созданию ряда акустических систем с линейно-фазовой характеристикой. Как известно из общей теории систем, для неискаженной передачи сигнала необходимо, чтобы соблюдались постоянство модуля передаточной функции, т.е. амплитудно-частотной характеристики (огibaющей амплитудного спектра), и линейная зависимость фазового спектра от частоты, т.е. $\varphi(\omega) = -\omega T$.

Действительно, если амплитудная огibaющая спектра сохраняется постоянной, то, как было сказано выше, искажений звукового сигнала при этом не должно происходить. Требования же к сохранению линейности фазы во всем диапазоне частот, как показали исследования Блаурта, оказались избыточными. Было установлено, что слух реагирует в первую очередь на скорость изменения фазы (т.е. ее производную по частоте), которая называется "групповое время задерживания ГВЗ": $\tau = -d\varphi(\omega)/d\omega$.

В результате многочисленных субъективных экспертиз были построены пороги слышимости искажений ГВЗ (т.е. величины отклонения $\Delta\tau$ от ее постоянного значения) для различных речевых, музыкальных и шумовых сигналов. Эти слуховые пороги зависят от частоты, и в области максимальной чувствительности слуха составляют 1...1,5 мс (рисунок 1). Поэтому последние годы при создании акустической аппаратуры Hi-Fi ориентируются, в основном, на приведенные выше слуховые пороги по искажению ГВЗ.

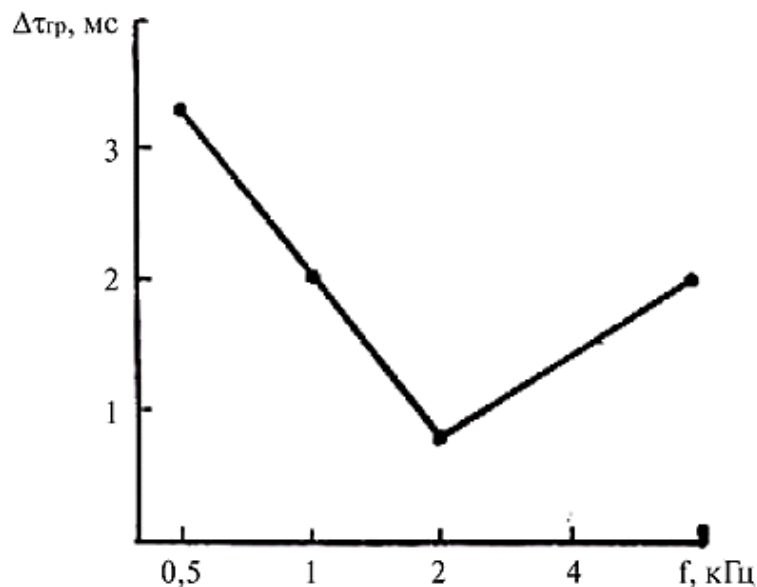


Рис. 1
Зависимость порогов слышимости искажений ГВЗ от частоты

Что касается влияния фазовых искажений на оценку тембра различных музыкальных инструментов, то были выполнены исследования на синтезированных звуках различных музыкальных инструментов по субъективным оценкам изменения тембра при введении различных фазовых искажений. В реальных музыкальных инструментах генерируются достаточно сложные сигналы с большим количеством обертонов, определенной негармоничностью между ними, сложным динамичным развитием во времени спектральной и временной огibaющих, а следовательно определенной динамикой развития во времени фазовых соотношений.

В одной из последних фундаментальных работ, посвященных этому вопросу (Галембо, Аскенфельд, Кадди - 2001 г.), были выполнены исследования на синтезированных звуках фортепиано. Для анализа был выбран низкочастотный диапазон (основные тона ниже 100 Гц), поскольку, как известно (см. "Звукорежиссер", 6/1999), при анализе сигнала в слуховой системе в области низких частот превалируют временные процессы.

Как уже было отмечено в той же статье, посвященной определению высоты тона, если какой-то из обертонов необычно ведет себя по амплитуде, то слуховая система выделяет его из общего ряда, и назначает ему отдельную высоту тона. Оказалось, что подобное поведение в фазовых спектрах также приводит к аналогичным результатам.

Кроме того, при оценке высоты тона "неразвернутых" гармоник, определение высоты происходит по их огibaющей, частота которой равна фундаментальной частоте. Если все обертоны музыкального тона находятся в фазе, то периодичность огibaющей становится четко выраженной, и возрастает точность определения высоты тона ("сила высоты тона"). Если фазовые соотношения между гармониками становятся

различными, то волновая структура суммарного звука претерпевает существенные изменения (пример показан на рисунке 2), и высота тона становится менее определенной (уменьшается сила высоты звука).

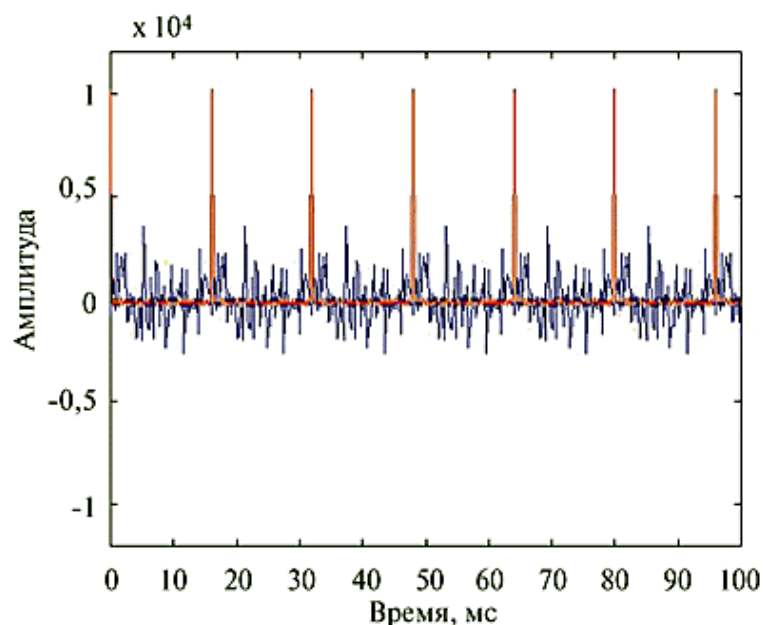


Рис. 2

Вид волновой формы при разных соотношениях фаз обертонов; красная - все обертоны имеют одинаковые начальные фазы, синяя - фазы распределены случайно

Таким образом, если фазовые соотношения оказывают слышимое влияние на определение высоты тона, то можно ожидать, что они окажут существенное влияние и на распознавание тембра.

Для экспериментов были выбраны звуки с основным тоном 27,5 и 55 Гц и со ста обертонами, с равномерным соотношением амплитуд, характерным для звуков фортепиано. При этом исследовались и тоны со строго гармоничными обертонами, и с определенной характерной для звуков фортепиано негармоничностью, которая возникает из-за конечной жесткости струн, их неоднородности, наличия продольных и крутильных колебаний и др.

Исследуемый звук синтезировался как сумма его обертонов: $X(t) = \sum A(n) \sin[2\pi f(n)t + \varphi(n, 0)]$

Для слуховых экспериментов были выбраны следующие соотношения начальных фаз для всех обертонов:

- А - синусоидальная фаза, начальная фаза была принята равной нулю для всех обертонов $\varphi(n, 0) = 0$;
- Б - альтернативная фаза (синусоидальная для четных и косинусоидальная для нечетных), начальная фаза $\varphi(n, 0) = \pi/4[(-1)^n + 1]$;
- С - случайное распределение фаз; начальные фазы при этом изменялись случайным образом в интервале от 0 до 2π .

В первой серии экспериментов все сто обертонов имели одинаковые амплитуды, различались только их фазы (основной тон 55 Гц). При этом прослушиваемые тембры получились различными:

- в первом случае (А), прослушивалась отчетливая периодичность;
- во- втором (Б), тембр был ярче и прослушивалась еще одна высота тона на октаву выше первой (правда высота не была четкой);
- в третьем (С) - тембр получился более равномерный.

Необходимо заметить - вторая высота прослушивалась только в наушниках, при прослушивании через громкоговорители все три сигнала отличались только тембром (сказывалась реверберация).

Это явление - изменение высоты тона при изменении фазы некоторых составляющих спектра - можно объяснить тем, что при аналитическом представлении преобразования Фурье сигнала типа Б, его можно представить как сумму двух комбинаций обертонов: сто обертонов с фазой типа А, и пятьдесят обертонов с фазой, отличающейся на $3\pi/4$, и амплитудой больше в $\sqrt{2}$. Этой группе обертонов слух назначает отдельную высоту тона. Кроме того, при переходе от соотношения фаз А к фазам типа В смещается центроид спектра (максимум энергии) в сторону высоких частот, поэтому тембр кажется ярче.

Аналогичные эксперименты со сдвигом фаз отдельных групп обертонов также приводят к появлению дополнительной (менее ясной) виртуальной высоты тона. Это свойство слуха связано с тем, что слух сравнивает звук с определенным имеющимся у него образцом музыкального тона, и если какие-то гармоники выпадают из типичного для данного образца ряда, то слух выделяет их отдельно, и назначает им отдельную

высоту.

Таким образом, результаты исследований Галембо, Аскенфельда и др. показали, что фазовые изменения в соотношениях отдельных обертонов достаточно отчетливо слышны как изменения тембра, и в некоторых случаях - высоты тона.

Особенно это проявляется при прослушивании реальных музыкальных тонов фортепиано, в которых амплитуды обертонов убывают с увеличением их номера, имеют особую форму огибающей спектра (формантной структуры), и отчетливо выраженная негармоничность спектра (т.е. сдвиг частот отдельных обертонов по отношению к гармоническому ряду).

Во временной области наличие негармоничности приводит к дисперсии, то есть высокочастотные компоненты распространяются по струне с большей скоростью, чем низкочастотные, и волновая форма сигнала изменяется. Наличие небольшой негармоничности в звуке (0,35%) добавляет некоторую теплоту, жизненность звучания, однако, если эта негармоничность становится большой, в звучании становятся слышны биения и другие искажения.

Негармоничность приводит также к тому, что если в начальный момент фазы обертонов находились в детерминированных соотношениях, то при ее наличии соотношения фаз со временем становятся случайными, пиковая структура волновой формы сглаживается, и тембр становится более равномерным - это зависит от степени негармоничности. Поэтому мгновенное измерение регулярности соотношения фаз между соседними обертонами может служить индикатором тембра.

Таким образом, эффект фазового перемешивания за счет негармоничности проявляется в некотором изменении восприятия высоты тона и тембра. Необходимо заметить, что эти эффекты слышны при прослушивании на близком расстоянии от деки (в позиции пианиста) и при близком расположении микрофона, причем слуховые эффекты различаются при прослушивании в наушниках и через громкоговорители. В реверберационном окружении сложный звук с высоким пик-фактором (что соответствует высокой степени регуляризации фазовых соотношений) говорит о близости источника звука, поскольку по мере удаления от него фазовые отношения приобретают все более случайный характер за счет отражений в помещении. Этот эффект может служить причиной разных оценок звучания пианистом и слушателем, а также разного тембра звука, записанного микрофоном у деки и у слушателя. Чем ближе, тем выше регуляризация фаз между обертонами и более определенная высота тона, чем дальше, тем более равномерный тембр и менее четкая высота.

Работы по оценке влияния фазовых соотношений на восприятие тембра музыкального звука сейчас активно изучаются в различных центрах (например, в ИРКАМе), и можно ожидать в ближайшее время новых результатов.

Тембр и общие принципы распознавания слуховых образов

Тембр является идентификатором физического механизма образования звука по ряду признаков, он позволяет выделить источник звука (инструмент или группу инструментов), и определить его физическую природу.

Это отражает общие принципы распознавания слуховых образов, в основе которых, как считает современная психоакустика, лежат принципы гештальт-психологии (geschalt, нем. - "образ"), которая утверждает, что для разделения и распознавания различной звуковой информации, приходящей к слуховой системе от разных источников в одно и то же время (игра оркестра, разговор многих собеседников и др.) слуховая система (как и зрительная) использует некоторые общие принципы:

- сегрегация - разделение на звуковые потоки, т.е. субъективное выделение определенной группы звуковых источников, например, при музыкальной полифонии слух может отслеживать развитие мелодии у отдельных инструментов;
- подобие - звуки, похожие по тембру, группируются вместе и приписываются одному источнику, например, звуки речи с близкой высотой основного тона и похожим тембром определяются, как принадлежащие одному собеседнику;
- непрерывность - слуховая система может интерполировать звук из единого потока через маскер, например, если в речевой или музыкальный поток вставить короткий отрезок шума, слуховая система может не заметить его, звуковой поток будет продолжать восприниматься как непрерывный;
- "общая судьба" - звуки, которые стартуют и останавливаются, а также изменяются по амплитуде или частоте в определенных пределах синхронно, приписываются одному источнику.

Таким образом, мозг производит группировку поступившей звуковой информации как последовательную, определяя распределение по времени звуковых компонент в рамках одного звукового потока, так и параллельную, выделяя частотные компоненты присутствующие и изменяющиеся одновременно. Кроме того, мозг все время проводит сравнение поступившей звуковой информации с "записанными" в процессе обучения в памяти звуковыми образами. Сравнивая поступившие сочетания звуковых потоков с имеющимися образами, он или легко их идентифицирует, если они совпадают с этими образами, или, в случае неполного совпадения,

приписывает им какие-то особые свойства (например, назначает виртуальную высоту тона, как в звучании колоколов).

Во всех этих процессах распознавание тембра играет принципиальную роль, поскольку тембр является механизмом, с помощью которого экстрактируются из физических свойств признаки, определяющие качество звука: они записываются в памяти, сравниваются с уже записанными, и затем идентифицируются в определенных зонах коры головного мозга (Рис. 3).

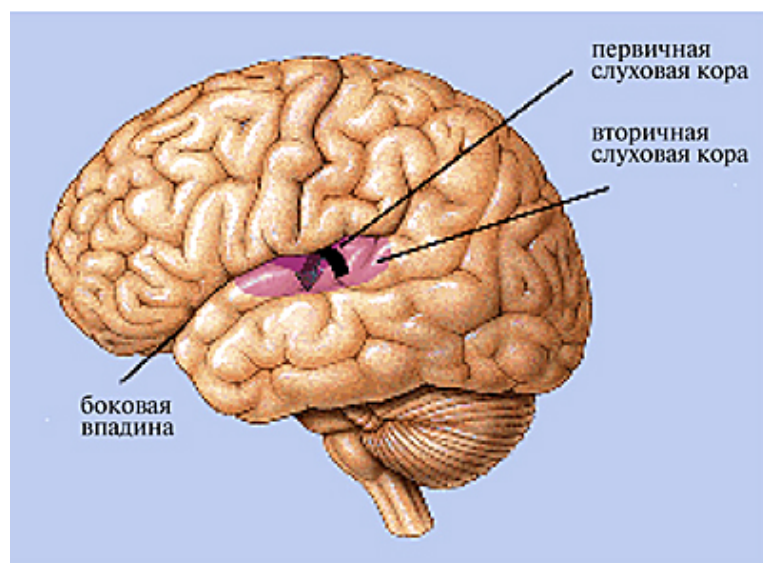


Рис. 3
Слуховые зоны мозга

Тембр - ощущение многомерное, зависящее от многих физических характеристик сигнала и окружающего пространства. Были проведены работы по шкалированию тембра в метрическом пространстве (шкалы - это различные спектрально-временные характеристики сигнала, см. вторую часть статьи в предыдущем номере). В последние годы, однако, появилось понимание, что классификация звуков в субъективно воспринимаемом пространстве не соответствует обычному ортогональному метрическому пространству, там происходит классификация по "субпространствам", связанным с вышеуказанными принципами, которые и не метрические, и не ортогональные.

Разделяя звуки по этим субпространствам, слуховая система определяет "качество звука", то есть тембр, и решает, к какой категории отнести эти звуки. Однако следует отметить, что все множество субпространств в субъективно воспринимаемом звуковом мире строится на основе информации о двух параметрах звука из внешнего мира - интенсивности и времени, а частота определяется временем прихода одинаковых значений интенсивности. Тот факт, что слух разделяет поступившую звуковую информацию сразу по нескольким субъективным субпространствам, повышает вероятность того, что в каком-то из них она может быть распознана. Именно на выделение этих субъективных субпространств, в которых происходит распознавание тембров и других признаков сигналов, и направлены усилия ученых в настоящее время.

Заключение

Подводя некоторые итоги, можно сказать, что основными физическими признаками, по которым определяется тембр инструмента, и его изменение во времени, являются:

- выстраивание амплитуд обертонов в период атаки;
- изменение фазовых соотношений между обертонами от детерминированных к случайным (в частности, за счет негармоничности обертонов реальных инструментов);
- изменение формы спектральной огибающей во времени во все периоды развития звука: атаки, стационарной части и спада;
- наличие нерегулярностей спектральной огибающей и положение спектрального центроида (максимума спектральной энергии, что связано с восприятием формант) и их изменение во времени (Рис. 4);

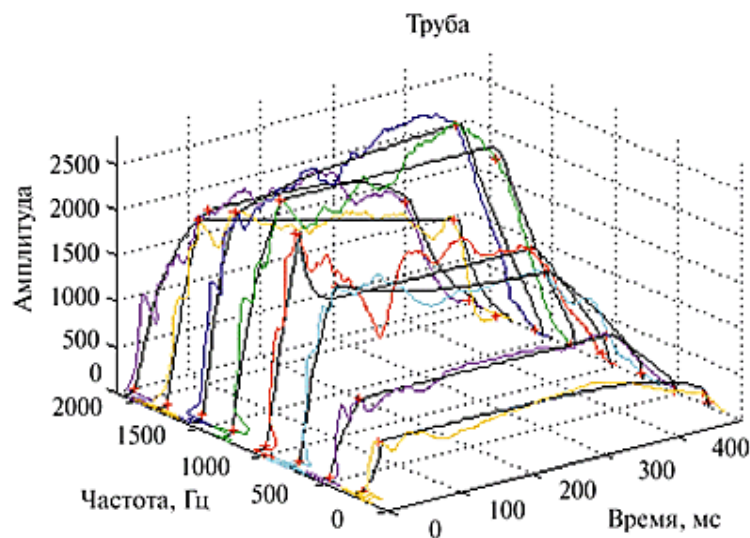


Рис. 4
Общий вид спектральных огибающих и их изменение во времени

- наличие модуляций - амплитудной (тремоло) и частотной (вibrато);
- изменение формы спектральной огибающей и характера ее изменения во времени;
- изменение интенсивности (громкости) звучания, т.е. характера нелинейности звукового источника;
- наличие дополнительных признаков идентификации инструмента, например, характерный шум смычка, стук клапанов, скрип винтов на рояле и др.

Разумеется, все это не исчерпывает перечень физических признаков сигнала, определяющих его тембр. Поиски в этом направлении продолжаются.

Однако при синтезе музыкальных звуков необходимо учитывать все признаки для создания реалистичного звучания.

Интересная классификация инструментов была предложена в IRCAME - "бинарное дерево" (рисунок 5). Если выделить пять признаков, и оценить их по 30-бальной шкале (негармоничность, форма атаки, форма спада, vibrато, тремоло, форма огибающей и др.), то все пять исследуемых инструментов можно расположить на "бинарном дереве", что может соответствовать их классификации по тембрам.

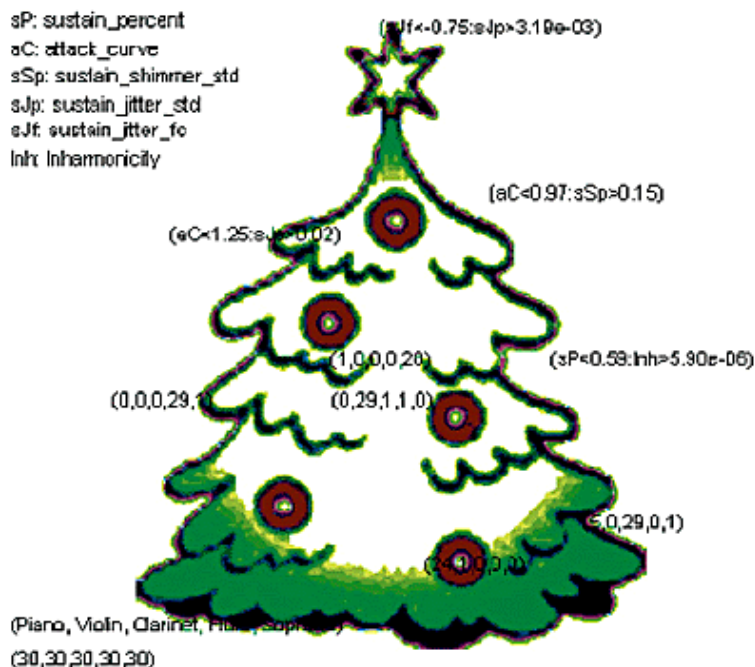


Рис. 5
"Бинарное дерево"

Приведенные в этих трех статьях сведения являются только несколькими первыми шагами в этом направлении, и далеко не исчерпывают проблемы. Надеюсь, что у нас будет возможность возвращаться к этим проблемам в дальнейшем, а также, надеюсь, что поставленные здесь вопросы заинтересуют наших читателей, и подтолкнут их к проведению научных и практических исследований в направлении анализа и восприятия тембра.

Приложение

Вербальное (словесное) описание тембра

Если для оценки высоты звуков имеются соответствующие единицы измерения: психофизические (мелы), музыкальные (октавы, тоны, полутоны, центы); есть единицы для громкости (соны, фоны), то для тембров такие шкалы построить невозможно, поскольку это понятие многомерное. Поэтому, наряду с описанными выше поисками корреляции восприятия тембра с объективными параметрами звука, для характеристики тембров музыкальных инструментов пользуются словесными описаниями, подобранными по признакам противоположности: яркий - тусклый, резкий - мягкий и др.

В научной литературе имеется большое количество понятий, связанных с оценкой тембров звука. Например, анализ терминов, принятых в современной технической литературе, позволил выявить наиболее часто встречающиеся термины, показанные в таблице. Были сделаны попытки выявить самые значимые среди них, и провести шкалирование тембра по противоположным признакам, а также связать словесное описание тембров с некоторыми акустическими параметрами (см. таблицу 6 в первой части этой статьи, "Звукорежиссер" 2/2001).

Таблица

Основные субъективные термины для описания тембра, используемые в современной международной технической литературе (статистический анализ 30 книг и журналов)

Acidlike - кислый	forceful - усиленный	muffled - заглушенный	sober - трезвый (рассудительный)
antique - старинный	frosty - морозный	mushy - пористый	soft - мягкий
arching - выпуклый	full - полный	mysterious - загадочный	solemn - торжественный
articulate - разборчивый	fuzzy - пушистый	nasal - носовой	solid - твердый
austere - суровый	gauzy - тонкий	neat - аккуратный	somber - мрачный
bite, biting - кусачий	gentle - нежный	neutral - нейтральный	sonorous - звучный
bland - вкрадчивый	ghostlike - призрачный	noble - благородный	steely - стальной
blaring - ревуший	glassy - стеклянный	nondescript - неопиcуемый	strained - натянутый
bleating - блеющий	glittering - блестящий	nostalgic - ностальгический	strident - скрипучий
breathy - дыхательный	gloomy - унылый	ominous - зловещий	stringent - стесненный
bright - яркий	grainy - зернистый	ordinary - ординарный	strong - сильный
brilliant - блестящий	grating - скрипучий	pale - бледный	stuffy - душный
brittle - подвижный	grave - серьезный	passionate - страстный	subdued - смягченный
buzzy - жужжащий	growly - рычащий	penetrating - проникающий	sultry - знойный
calm - спокойный	hard - жесткий	piercing - пронзительный	sweet - сладкий
carrying - полетный	harsh - грубый	pinched - ограниченный	tangy - запутанный
centered - концентрированный	haunting - преследующий	placid - безмятежный	tart - кислый
clangorous - звенящий	hazy - смутный	plaintive - заунывный	tearing - неистовый
clear, clarity - ясный	heartly - искренний	ponderous - увесистый	tender - нежный
cloudy - туманный	heavy - тяжелый	powerful - мощный	tense - напряженный
coarse - грубый	heroic - героический	prominent - выдающийся	thick - толстый
cold - холодный	hoarse - хриплый	pungent - едкий	thin - тонкий
colorful - красочный	hollow - пустой	pure - чистый	threatening - угрожающий
colorless - бесцветный	honking - гудящий (автомобильный гудок)	radiant - сияющий	throaty - хриплый
cool - прохладный	hooty - гудящий	raspy - дребезжащий	tragic - трагичный
crackling - трескучий	husky - сиплый	rattling - грохочущий	tranquil - успокаивающий
crashing - ломаный	incandescence - накалиeнный	reedy - пронзительный	transparent - прозрачный
creamy - сливочный	incisive - резкий	refined - рафинированный	triumphant - торжествующий
crystalline - кристаллический	inexpressive - невыразительный	remote - удаленный	tubby - бочкообразный
cutting - резкий	intense - интенсивный	rich - богатый	turbid - мутный

dark - темный	introspective - углубленный	ringing - звенящий	turgid - высокопарный
deep - глубокий	joyous - радостный	robust - грубый	unfocussed - несфокусированный
delicate - деликатный	languishing - печальный	rough - терпкий	unobtrusive - скромный
dense - плотный	light - светлый	rounded - круглый	veiled - завуалированный
diffuse - рассеянный	limpid - прозрачный	sandy - песочный	velvety - бархатистый
dismal - отдаленный	liquid - водянистый	savage - дикий	vibrant - вибрирующий
distant - отчетливый	loud - громкий	screamy - кричащий	vital - жизненный
dreamy - мечтательный	luminous - блестящий	sere - сухой	voluptuous - пышный(роскошный)
dry - сухой	lush (luscious) - сочный	serene, serenity - спокойный	wan - тусклый
dull - скучный	lyrical - лирический	shadowy - затененный	warm - теплый
earnest - серьезный	massive - массивный	sharp - резкий	watery - водянистый
ecstatic - экстатический	meditative - созерцательный	shimmer - дрожащий	weak - слабый
ethereal - эфирный	melancholy - меланхолический	shouting - кричащий	weighty - тяжеловесный
exotic - экзотический	mellow - мягкий	shrill - пронзительный	white - белый
expressive - выразительный	melodious - мелодичный	silky - шелковистый	windy - ветряный
fat - жирный	menacing - угрожающий	silvery - серебристый	wispy - тонкий
fierce - жесткий	metallic - металлический	singing - певучий	woody - деревянный
flabby - дряблый	misty - неясный	sinister - зловещий	yearning - тоскливый
focussed - сфокусированный	mournful - траурный	slack - расхлябанный	
forboding - отталкивающий	muddy - грязный	smooth - гладкий	

Однако, главная проблема состоит в том, что нет однозначного понимания различных субъективных терминов, описывающих тембр. Приведенный в таблице перевод далеко не всегда соответствует тому техническому смыслу, которое вкладывается в каждое слово при описании различных аспектов оценки тембра.

В нашей литературе раньше был стандарт на основные термины, но сейчас дела обстоят совсем печально, поскольку не ведется работа по созданию соответствующей русскоязычной терминологии, и употребляется много терминов в разных, иногда прямо противоположных, значениях.

В связи с этим AES при разработке серии стандартов по субъективным оценкам качества аудиоаппаратуры, систем звукозаписи и др. начал приводить определения субъективных терминов в приложениях к стандартам, а так как стандарты создаются в рабочих группах, включающих ведущих специалистов разных стран, то эта очень важная процедура приводит к согласованному пониманию основных терминов для описания тембров.

В качестве примера приведу стандарт AES-20-96 - "Рекомендации для субъективной оценки громкоговорителей", - где дано согласованное определение таких терминов, как "открытость", "прозрачность", "ясность", "напряженность", "резкость" и др.

Если эта работа будет систематически продолжаться, то, возможно, основные термины для словесного описания тембров звуков различных инструментов и других звуковых источников будут иметь согласованные определения, и будут однозначно или достаточно близко пониматься специалистами разных стран. Мы постараемся информировать об этом наших читателей.

Часть 15. Слуховое восприятие пространственных систем. Часть 1

Последнее десятилетие характеризуется бурным развитием систем пространственной звукопередачи (недаром 109-й конгресс AES в Лос-Анжелесе назывался Surrounded by Sound). Разработка и широкое использование таких систем в значительной степени изменило технологию звукозаписи, принципы проектирования систем звуковоспроизведения и пр. Дальнейшее их развитие требует новой, более глубокой информации о различных аспектах пространственного слуха, без получения которой невозможно решать такие глобальные задачи, как проблемы переноса пространственного звукового образа из первичного помещения (концертного зала, студии и др.) во вторичное помещение прослушивания.

В статье "Бинауральный слух и пространственная локализация" ("Звукорежиссер" 10/1999) были приведены основные сведения, относящиеся к пространственной локализации единичного звукового источника в условиях отсутствия отражений (например, в безэховой камере или в большом хорошо заглушенном помещении). Эти данные уже достаточно широко известны в психоакустике и подробно освещены в литературе (например, в книге Блауэрта "Пространственный слух").

Однако для развития пространственных звуковых систем этой информации явно недостаточно. Поэтому в последние годы в разных странах проводятся многочисленные исследования по углубленному изучению возможностей слуховой системы в воссоздании пространственного звукового образа и оценке его тембральных характеристик. Результаты этих работ представлены в многочисленных докладах практически на всех последних конгрессах AES, на специальных конференциях, в статьях в таких журналах, как JAES, JASA, Acoustica и др.

Поскольку эти вопросы чрезвычайно важны для работы звукорежиссеров, инженеров, музыкантов и др, особенно в связи с освоением новых технологий записи и воспроизведения звука в системах Surround Sound, то представляется полезным рассказать о некоторых новых результатах, полученных в этом направлении за последнее время. Расскажу в основном об исследованиях, проводимых под руководством всемирно известных психоакустиков Б. Мура в Кембридже (Англия), Б Хартмана (США), а также о некоторых других исследованиях.

Большое практическое значение для развития пространственных аудиосистем имеют ответы на следующие вопросы:

- как слуховая система локализует звук в условиях, когда имеется много звуковых источников, работающих одновременно;
- как отражения в помещении влияют на пространственную локализацию;
- как в слуховой системе реализуется процесс локализации вообще.

Как уже было показано в вышеупомянутой статье, при локализации единичного источника точность локализации в горизонтальной плоскости достаточно высока и составляет примерно 3° , хотя имеются данные, что минимальное разрешение смещения источника может составлять даже 1° . Существуют два механизма локализации источника в горизонтальной плоскости:

- на низких частотах (до 1500 Гц) – это оценка разности по времени прихода звука к разным ушам (ITD). Интересно, что при угловом смещении в 1° разница по времени составляет ~ 10 мс, что показывает очень высокую точность оценки в слуховой системе (рисунок 1);
- на высоких частотах (выше 2 кГц) – это оценка разности по интенсивности ILD, возникающая за счет дифракции звука вокруг головы (рисунок 2).

Локализация в вертикальной плоскости происходит за счет модификации спектра источника звука при взаимодействии с ушной раковиной, головой и торсом, которые действуют как частотно-зависимые дифракционные фильтры.

Анализ работы этих фильтров все время продолжается и уточняется, поскольку это очень важно для развития

пространственных систем звуковоспроизведения и бинауральной стереофонии. Изменение формы АЧХ, измеренной в слуховом проходе, при перемещении единичного источника звука с равномерной АЧХ в вертикальной плоскости за счет дифракции на голове и ушной раковине показано на рисунке 3.

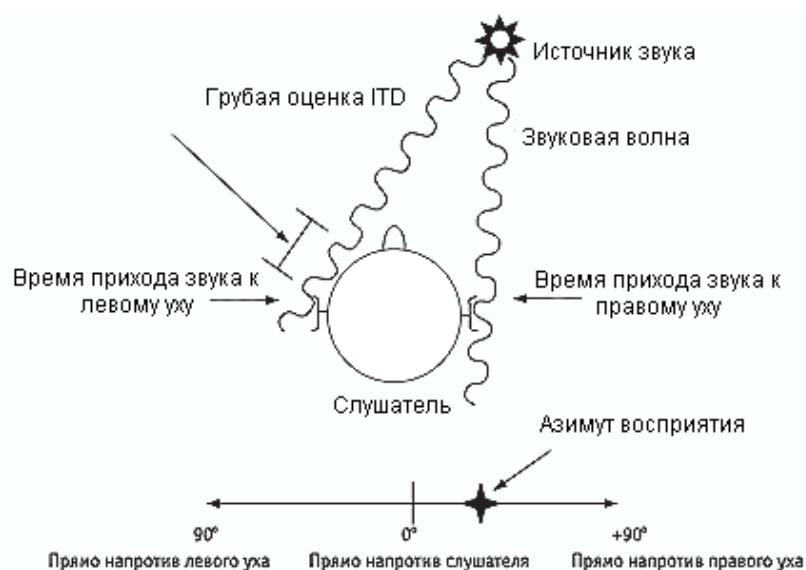


Рис.1 Локализация за счет разности во времени прихода звука-ITD

В целом точность локализации в вертикальной плоскости гораздо хуже, чем в горизонтальной, и составляет 19-20°, хотя минимально определяемый сдвиг составляет 4°, т.е. всего в четыре раза хуже, чем в горизонтальной плоскости.

Для ответа на первый поставленный выше вопрос – "Как происходит локализация, когда одновременно звучат несколько источников?", – были проведены многочисленные эксперименты, которые дали неутешительный ответ: точность локализации становится значительно хуже, даже если работают только два источника с частично перекрывающимися спектрами – именно это и имеет место в стереосистемах, пространственных системах воспроизведения и др.

Например, были поставлены эксперименты со слушателями по локализации в заглушенной камере, где по периметру размещались источники звука (рисунок 4). Более "гуманные" опыты были выполнены с помощью "искусственной головы", которая размещалась в заглушенной камере, где по периметру в горизонтальной плоскости на расстоянии 4 м были установлены громкоговорители. На "искусственной голове" записывались импульсные характеристики в слуховом проходе. Затем звуки, предварительно обработанные в соответствии с записанными на "искусственной" голове импульсными характеристиками, подавались на стереотелефоны, и слушателей просили локализовать два одновременно звучащих источника, т.е. определить их угловое расстояние. Измерения были выполнены для разных сигналов: пары синусоид, различных видов модулированных сигналов и т.д. Если спектры сигналов, подаваемых из разных источников, перекрывались мало, то точность локализации составляла 18°, если спектры перекрывались, то 60°. Проще говоря, чтобы можно было определить, что звуки идут от разных источников, они должны были быть разнесены на 60°.

Это намного хуже, чем локализация одиночного источника. Кажется, что эти результаты противоречат практике – мы постоянно слышим несколько источников одновременно и обычно достаточно хорошо их локализуем. Однако реальные звуковые сигналы (музыка, речь, шум) обладают некоторыми особенностями, которые помогают слуховой системе.

Трудности, которые возникают при локализации нескольких источников, связаны, во-первых, с тем, что звуки от нескольких источников складываются в обоих ушах, и становится трудно определить разницу по времени и по интенсивности для каждого источника отдельно, чтобы установить локализацию каждого из них. Это можно сделать, если спектры звуков от каждого источника существенно отличаются (например, находятся в разных частях диапазона) – тогда слуховая система выполняет спектральный анализ, и в этом случае она справляется с задачей локализации.

Во-вторых, проблема состоит том, что перед слуховой системой встают две разные и конкурирующие между собой задачи: и локализовать звуки, и определить, к какому источнику они принадлежат (т.е. выполнить их сегрегацию, разделение на звуковые потоки).

В моей статье в 4/2001 уже было немного сказано о механизме сегрегации, сейчас поговорим об этом подробнее.

Одними из очень важных критериев объединения звуков в один звуковой поток, то есть приписывания их

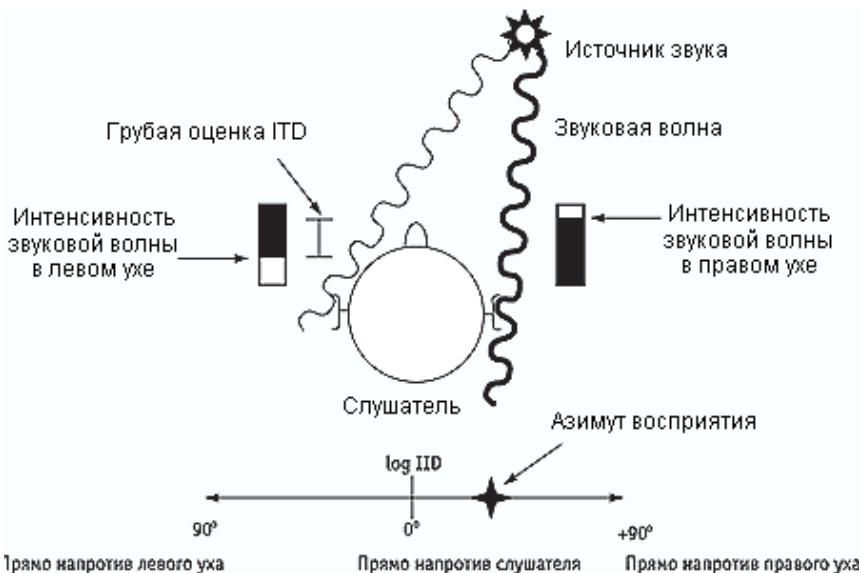


Рис.2 Локализация за счет разности интенсивностей-ITD

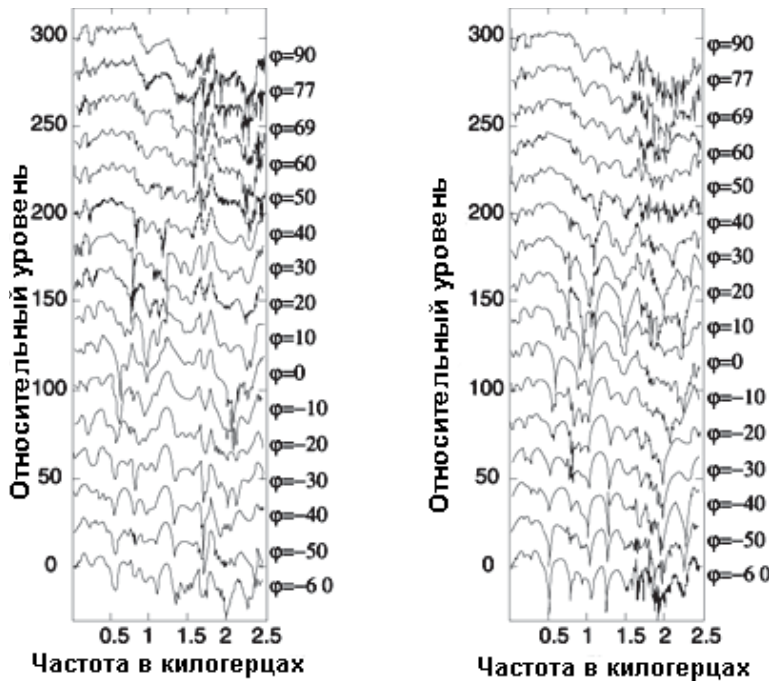


Рис.3 Форма АЧХ, измеренная в левом и правом слуховых каналах при разных углах подъема звукового источника

одному источнику, являются подобие спектров и характер переходных процессов, а также их синхронизация по времени – если звуки включаются и выключаются одновременно, то слуховая система обычно считает, что они исходят от одного источника, даже если на самом деле они разнесены в пространстве. Это очень важный вывод для пространственной звукозаписи. Реальные источники редко включаются и выключаются синхронно, кроме того их спектры постоянно меняются во времени – какие-то источники (инструменты) доминируют в разные моменты времени, поэтому слух успевает их выделить и локализовать.

В связи с этим возникают чрезвычайно интересные вопросы: "Зависит ли процесс сегрегации от процессов локализации звуков?" "Какой из процессов происходит сначала?" "Используются ли в процессе сегрегации какие-то признаки, по которым слуховая система локализует звуки?"

Обычно в процессе прослушивания звуков от различных источников (например, инструментов в оркестре) не возникает особых трудностей выделения отдельных источников звука. Насколько хорошо это делается, зависит от состояния слуха и степени музыкальной подготовленности слушателя. Когда в слуховую систему поступают звуки от двух звуковых источников, например от скрипки и рояля, то в высших отделах коры головного мозга (а не в периферической слуховой системе) происходит разделение всех поступивших звуков на два отдельных потока – один относится к роялю, другой – к скрипке. При этом происходят два разных процесса: один – разделение одновременно поступающих звуков, другой – последовательное разделение их во времени. Эти два процесса называются последовательной и параллельной группировкой (сегрегацией).

Для такого разделения используется большое количество различных физических признаков (часть из них была упомянута выше).

Эксперименты показали, что некоторые из признаков, по которым система производит локализацию в пространстве, (например разница во времени прихода звуков к разным ушам – ITD), оказывают



Рис.4 Размещение громкоговорителей на сфере в заглушенной камере

существенное влияние и на процесс последовательной группировки – но оказывают относительно малое влияние на процесс одновременной группировки. Это подтверждается различными экспериментами: например, через головные телефоны одновременно предъявлялись разные гласные звуки, и слушателей просили их различить. Если между звуками вводилась задержка во времени 400 мс (что соответствует углу между ними в 45° , как если бы они воспроизводились через разные громкоговорители), то различимость улучшалась всего на 7%, в то же время разница в полтона по основной частоте улучшала их различимость на 22%. Однако при предъявлении последовательных звуков введение дополнительной временной задержки (ITD) существенно улучшало их разделимость.

Результаты исследований позволяют сделать важный для практики вывод: при прослушивании одновременных звуков от разных источников слуховая система сначала производит их группировку по потокам (определяет, какие звуки к какому источнику принадлежат), используя при этом различные непространственные признаки (значение основной частоты, степень гармоничности, амплитудную огибающую, структуру переходных процессов и др), а затем локализует данный источник звука, т. е. сначала определяет, "Что это", а затем – "Где это"? (рисунок 5).

Затем слух продолжает разделять последовательно поступающие звуки по потокам (каждому инструменту – свой), но при этом уже использует пространственные признаки: разницу во времени поступления, разницу в амплитудах в разных ушах и др. Интересно, что если при этом источник звука не меняет своего положения или изменяет его плавно (например, солист на сцене), то слуховая система продолжает воспринимать этот звук как часть одного слухового потока, то есть считает исходящим от единого звукового источника. Но если источник звука резко меняет свое положение в пространстве (локализацию), то слуховая система может воспринять его как совершенно другой источник звука – процесс сегрегации произведет расщепление звукового потока.

Учитывая сложность одновременно действующих процессов при восприятии нескольких звуковых источников, разнесенных в пространстве, например, в системах Surround Sound, неудивительно, что точность локализации при этом ухудшается.

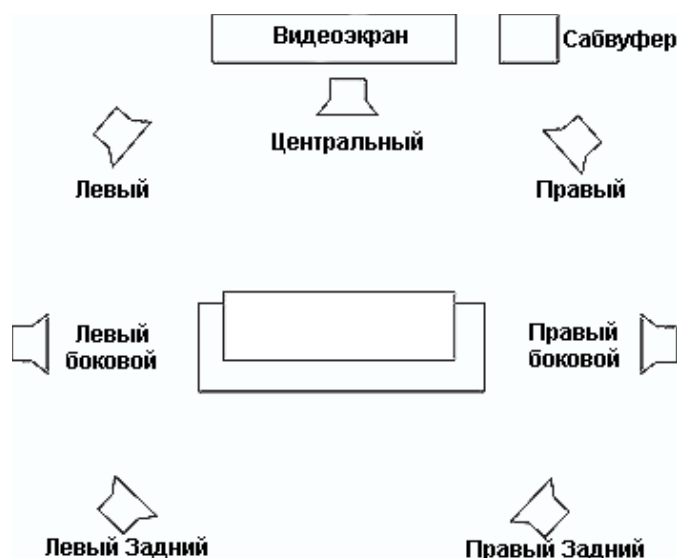


Рис.5 Процесс сегрегации и локализации

Следующая проблема, которая возникает при локализации нескольких источников звука, состоит в оценке

влияния на точность локализации отраженных сигналов, которые возникают в любом помещении прослушивания (студии, концертном зале, комнате прослушивания). Как известно (см. "Звукорежиссер" 10/2000), структура отраженных сигналов в помещении имеет вид, представленный на рисунке 6. Если источник звука излучает короткий импульс, то к слушателю (или микрофону), находящемуся в определенном месте помещения, сначала приходит прямой сигнал, затем, через определенные отрезки времени, первые отраженные сигналы (как правило, от потолка, боковых стен, пола), затем количество этих отраженных сигналов возрастает, и процесс спада уровня сигнала становится почти непрерывным. Время, в течение которого сигнал спадает на 60 дБ, называется "временем стандартной реверберации".

Процесс реверберации определяет качество тембра воспринимаемого звука в помещении, кроме того, он оказывает существенное влияние на процесс локализации источников звука в помещении. Наличие отраженных звуков, приходящих со всех направлений, создает случайные вариации признаков, определяющих локализацию, и неизбежно ухудшает ее точность. Особенно страдает от этого такой признак локализации, как разница во времени между сигналами на двух ушах (ITD). В помещении, где отраженные звуки преобладают над прямыми, этот критерий локализации вообще становится ненадежным. Зато второй критерий, разница по интенсивности (IID), "страдает" меньше, так как он используется на высоких частотах, где коэффициент поглощения на поверхности помещения значительно возрастает с частотой, и уровень отраженных сигналов уменьшается. Если локализуются широкополосные сигналы от разных источников в сильно реверберирующем помещении, то слух, в основном, полагается на информацию от высокочастотной части спектра, используя только второй критерий (IID). Такая переоценка локализационных признаков происходит на подсознательном уровне.

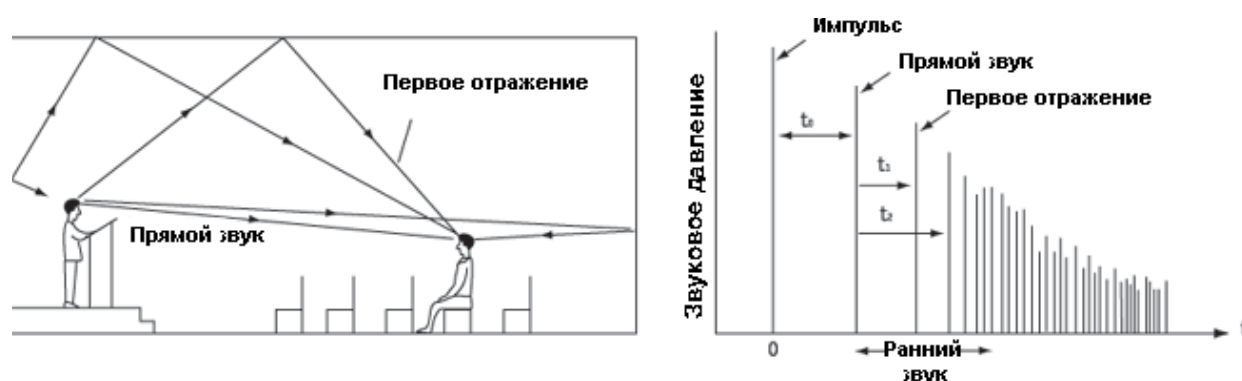


Рис.6 Структура отраженных сигналов в помещении

Второй механизм, помогающий осуществлять локализацию источников звука в помещении, — "эффект предшествования", или "эффект Хааса", или "закон первой волны". Явление это известно достаточно давно, однако его объяснение с точки зрения современной психоакустики появилось только в настоящее время.

Сущность этого явления заключается в следующем: если звуки с коротким интервалом задержки по времени приходят с разных направлений, то локализация общего звука происходит по первому пришедшему звуку, т.е. слуховая система как бы теряет способность локализовать отраженный звук, если он приходит через слишком короткий отрезок времени (рисунок 7).

Слуховая система отдает предпочтение первому "прямому" звуку, который несет более точные данные о локализации источника по сравнению с отраженным звуком, который искажает информацию о локализации. Это своего рода "нейронные ворота", которые открываются в момент атаки звука, производят его локализацию и закрываются.

Нужно сказать, что этот отраженный звук все-таки оказывает свое влияние на точность локализации первого основного звука — если место появления отраженного звука все больше отодвигается от расположения прямого звука, то он как бы "утягивает" локализацию первого звука за собой примерно на 7° (меняется протяженность первого источника); при большем смещении эффект уже не сказывается.

Если интервал между двумя короткими звуками становится слишком коротким (меньше 1 мс), то эффект предшествования не проявляется, происходит некоторая компромиссная (усредненная) локализация. Этот эффект называется "суммарная локализация". Если интервал больше 5 мс для импульсов (щелчков) и больше 40 мс для речи и музыки, то слышны отдельно и прямой звук, и эхо, то есть эффект предшествования также не проявляется.

Если уровень отраженного звука сделать на 10...15 дБ больше уровня прямого звука, то эффект предшествования также пропадает – слышны два разных звука с разных направлений.

Этот эффект проявляется обычно, когда два звука похожи по спектру, что и имеет место в прямом и отраженном сигнале. Однако эксперименты показали, что эффект имеет место и для двух разных звуков, например, прямой низкочастотный сигнал очень эффективно подавляет локализацию высокочастотного отраженного сигнала.

Интересно отметить, что этот эффект проявляется не только тогда, когда сигналы приходят из разных направлений в горизонтальной плоскости, где основную роль играет разница по времени и интенсивности. Эффект Хааса имеет место и при локализации прямого и отраженного звуков в вертикальной плоскости, правда, он выражен значительно слабее.

Необходимо сказать, что эффект предшествования не подавляет всю информацию об отраженном сигнале – слушатель легко различает разницу в тембрах прямого звука и звука, дополненного отражениями. Эта тембральная разница несет информацию о размерах помещения, позиции стен, потолка и др. Эффект предшествования проявляется только в том, что ранние отраженные сигналы не слышны как отдельные звуки, и информация об их пространственной локализации теряется.

Таким образом, точность локализации источников звука в помещении при наличии отражений существенно ухудшается по обычным критериям (ITD, IID). Однако слух использует два других механизма – локализацию по IID в высокочастотной части диапазона и эффект предшествования, что позволяет осуществлять локализацию, хотя и с меньшей точностью.

Одни из самых последних исследований в психоакустике посвящены третьей проблеме: "Как в слуховой системе реализуется процесс подавления первых отражений и процесс локализации вообще?". Является ли это следствием работы специализированных нейронов, или это продукт сознательного принятия решений высшими отделами головного мозга?

Исследования на животных позволили выявить специальные бинауральные нейроны, способные сравнивать сигналы от двух ушей и реагировать на разницу во времени и на разницу в интенсивности между ними. На нейронах в этих же отделах мозга было выявлено, что при подаче двух коротких щелчков с изменяемой задержкой между ними чувствительность нейронов ко второму звуку при коротких задержках подавляется. Что касается реакции человека, то, если бы эффект предшествования определялся только реакцией нейронов, он происходил бы практически мгновенно. Однако выяснилось, что он требует определенного времени для возникновения, то есть слуховая система как бы "обучается".

Например, были проделаны такие эксперименты: если подать два сигнала с задержкой 8 мс (что моделирует как бы прямой звук и его задержанное эхо), то в первый момент эти два сигнала слышны раздельно, но если их повторить несколько раз, например со скоростью четыре раза в секунду, то через некоторое время второй звук перестает быть слышимым. Эффект предшествования может быть разрушен резким изменением акустической обстановки: если один сигнал подавать от одного громкоговорителя, а другой с некоторой задержкой от другого, то после определенного периода обучения возникает эффект подавления, но, если внезапно изменить расположение громкоговорителей (или одного из них), то эффект пропадает, и каждый звук слышен отдельно.

Все эти эксперименты заставляют предположить, что восприятие эффекта предшествования является актом сознания, а не физиологической особенностью. Создается впечатление, что эффект предшествования срабатывает только тогда, когда время появления эха, его амплитуда и направление совпадают с некоторым "ожиданием" слушателя от акустики данного помещения. Это ожидание формируется на основании предшествующего опыта прослушивания в данном зале (или подобных), зрительного впечатления, предварительного обучения и др. Однако быстрое изменение позиции ведущего и ведомого звука, несовпадающего со слушательскими ожиданиями от акустики данного помещения, сразу же делают эхо слышимым, то есть нарушают эффект. Аналогичные результаты получаются при изменении спектра эха или направления его прихода, которые, по мнению слушателя, делают его неестественным для данного

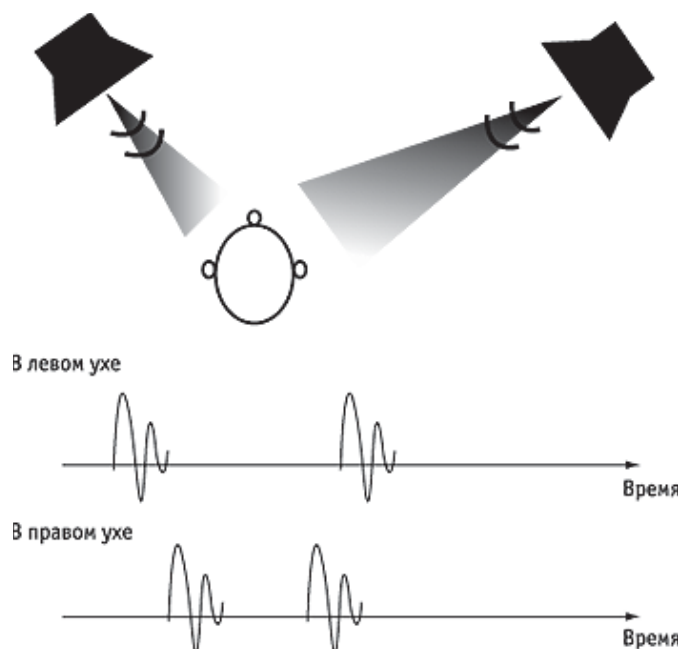


Рис.7 Эффект предшествования

помещения, что также делает его слышимым.

Таким образом, как только нарушаются траектории прихода звука и его параметры, выстроенные слушателем в сознании при предварительном анализе акустики данного помещения, так эффект предшествования сразу пропадает.

Создается впечатление, что сначала прямой звук и эхо-сигнал обрабатываются слуховой системой совместно, обогащая спектр (тембр) прямого звука, затем оба сигнала обрабатываются высшей нервной системой, и она принимает решение – подходит ли данный звук по своим параметрам на роль эха от прямого сигнала в данном помещении. Если подходит, то информация о нем подавляется, и локализация происходит только по прямому звуку; если не подходит, то он слышен как отдельный звук, и происходит локализация двух разных источников.

Следовательно, при локализации звуков от разных, одновременно работающих источников в реальном помещении (например, в системах домашнего театра) слуховой системе в процессе локализации, то есть в процессе построения пространственного образа, приходится решать сразу две различные и трудные задачи:

- произвести классификацию всех поступивших звуков по потокам, определить к какому источнику какой звук принадлежит;
- произвести локализацию всех источников, для чего необходимо еще решить проблему, какие звуки идут прямо от источников, а какие являются их отраженными сигналами (эхо), и поэтому их можно не принимать во внимание при локализации.

Неудивительно, что при таких условиях общая точность локализации снижается, и, соответственно, увеличивается время ее реализации.

Восприятие пространственных звуковых систем зависит не только от точности локализации, но и от особенностей восприятия тембра пространственного слухового образа. Об этом речь пойдет в следующей части этой статьи.

Часть 15. Слуховое восприятие пространственных систем, часть 2

Развитие систем стереовоспроизведения и современных систем пространственного звуковоспроизведения (Home Theatre и др.) основано на создании слуховых иллюзий – большом обмане слуховой системы. Две или большее число акустических систем (например, 5.1) воспроизводят звук из разных точек помещения, а слушатель слышит звуки, исходящие из несуществующих (мнимых) источников, воспринимая определенную звуковую панораму. Скорее всего, этот обман оказывается возможным потому, что все системы звуковоспроизведения являются искусственными, сравнительно недавно созданными устройствами, поэтому слух в процессе длительного биологического развития оказался неприспособленным для распознавания этого обмана (возможно, существует и более глубокое объяснение этого явления).

Однако бурное развитие пространственных систем звуковоспроизведения и многочисленные психоакустические эксперименты показали, что обман происходит не полностью: слух действительно локализует мнимый источник там, где никакого реального источника не существует, – но вот слышит ли он тембр от этого источника таким же, как от реального источника, находящегося на этом же месте, – это очень большой вопрос.

В рамках большого исследовательского международного проекта Medusa по исследованию восприятия пространственных систем звуковоспроизведения были поставлены работы в исследовательских центрах различных стран (Англии, Дании, США и др.) по анализу тембров мнимых источников в разных условиях звуковоспроизведения. Исследования эти особенно актуальны в настоящее время не только в связи с развитием пространственных систем, но и созданием виртуального акустического пространства (например, систем аурализации, "Звукорежиссер", 7/2000). Результаты, полученные на первых этапах этих исследований, могут быть полезны звукорежиссерам, осваивающим новые технологии пространственной звукозаписи.

Как уже было показано в статьях, посвященных тембру ("Звукорежиссер", 2, 3/2001), тембр определяется в соответствии с международными стандартами как "атрибут слухового ощущения, с помощью которого слушатель может судить, что два звука, одинаковые по высоте и громкости, отличаются друг от друга". Оценка тембра зависит от многих факторов: амплитудного спектра и его изменения во времени, фазового спектра, процессов установления и спада звука и др.

На начальном этапе исследований тембров виртуальных источников и их отличий от реальных источников были проведены работы по сравнительной оценке только амплитудных спектров. Было высказано предположение, что для оценки тембров используется формирование спектральной плотности, которое лежит в основе определения громкости, а именно – распределение спектральных уровней внутри третьоктавных полос. Распределение третьоктавных полос очень близко к частотно-зависимой ширине критических полос слуха (рисунок 1). Под "критической полосой" ("Звукорежиссер", 9/2000) понимается ширина полосы пропускания слуховых фильтров, с помощью которых происходит спектральный анализ звуковых сигналов на базилярной мембране во внутреннем ухе.

Ширина этих полос зависит от частоты и интенсивности сигнала. Наличие критических полос представляет собой очень важное свойство слуховой системы. Внутри каждой критической полосы происходит интеграция энергии независимо от вида сигнала: отрезок шума или спектральные компоненты тональных сигналов, находящиеся внутри критической полосы, если они имеют одинаковый уровень интенсивности, интегрируются и создают ощущение одинакового уровня удельной громкости. Поэтому общее ощущение громкости связано с распределением спектрального уровня сигнала по критическим полосам и определяется возникающими при этом ощущениями удельной громкости.

В работах Цвиккера было показано, что бинауральные эффекты при восприятии тембров зависят от сумм спектральных удельных громкостей, поступающих из каждого ушного канала в центральные отделы слуховой системы.

Компьютерная модель периферической слуховой системы, построенная для оценки бинауральной громкости, была использована и для оценки различия тембров реальных и виртуальных источников. Принцип работы компьютерной модели слуховой системы, показанный на рисунке 2, использует построение передаточных функций левого и правого уха (HRTF) с учетом дифракции на голове и ушной раковине, и фильтрации сигнала во внешнем и среднем ухе. Кроме того, в модели происходит спектральная обработка сигнала с помощью системы полосовых фильтров в 42 полосах, характеристики которых аналогичны слуховым фильтрам на базилярной мембране. Затем подсчитывается энергия сигнала на выходе каждого фильтра, и из него определяется удельная спектральная громкость. Полученное таким образом спектральное распределение удельной громкости и используется в дальнейших экспериментах для оценки тембров мнимых источников.

Для оценки параметров виртуального источника (его локализации и тембра) были использованы различные конфигурации расположения громкоговорителей в помещении: два громкоговорителя, расположенные на равном расстоянии от слушателя (обычная стереопара, рисунок 3); три громкоговорителя, расположенные в виде треугольника; громкоговорители, расположенные в различной пространственной позиции, в том числе в вертикальной плоскости. Кроме того, оценки выполнялись при поворотах головы слушателя (рисунок 3а, 3б). На громкоговорители подавался когерентный (одинаковый) сигнал, панорамирование (создание и размещение виртуального источника) происходило за счет изменения амплитуд сигнала в каждом канале громкоговорителя путем изменения коэффициента усиления.

Как было указано в первой части этой статьи, локализация мнимого источника при расположении двух громкоговорителей в горизонтальной плоскости происходит за счет разницы сигналов во времени (ITD) и разницы по интенсивности (IID) между сигналами, попадающими на разные уши. Как показано на рисунке 3, определение направления на источник звука в горизонтальной плоскости производится с помощью расчета угла θ_{cc} . Для расположения двух громкоговорителей в горизонтальной плоскости угол на мнимый источник при интенсивностном (амплитудном) панорамировании (только за счет изменения амплитуд сигналов от громкоговорителей) определяется по формуле:

$$\frac{\lg \theta_{cc}}{\lg \theta_0} = \frac{g_1 - g_2}{g_1 + g_2}$$

где θ_{cc} и θ_0 – углы на мнимый источник и на реальный громкоговоритель, g_1 и g_2 – коэффициенты усиления сигналов в громкоговорителях. Этот широко известный в акустике закон дает достаточную точность при определении направления на мнимый источник для стереопары громкоговорителей.

Для произвольно расположенных в трехмерном пространстве громкоговорителей направление на мнимый источник определяется с помощью вектора пространственного панорамирования P_T , который выражается в матричной форме: $P_T = g L_{nmk}$, где L_{nmk} – вектор направлений на громкоговорители (рисунок 4), P_T – вектор направлений на мнимые источники, g – вектор коэффициентов усиления. Этот метод определения положения мнимых источников (VBAP) был предложен сравнительно недавно, в 1997 году финским ученым В. Пулки. С его использованием были созданы цифровые процессоры, позволяющие управлять в реальном времени положением мнимых источников в трехмерном пространстве (рисунок 5). В частности, было реализовано цифровое устройство на восемь входов и восемь выходов, то есть восемь громкоговорителей, расположенных в пространстве, которые были успешно использованы в системах мультимедиа и даже в больших залах на концертах компьютерной музыки. Возможно, сейчас имеются устройства и для большего числа каналов.

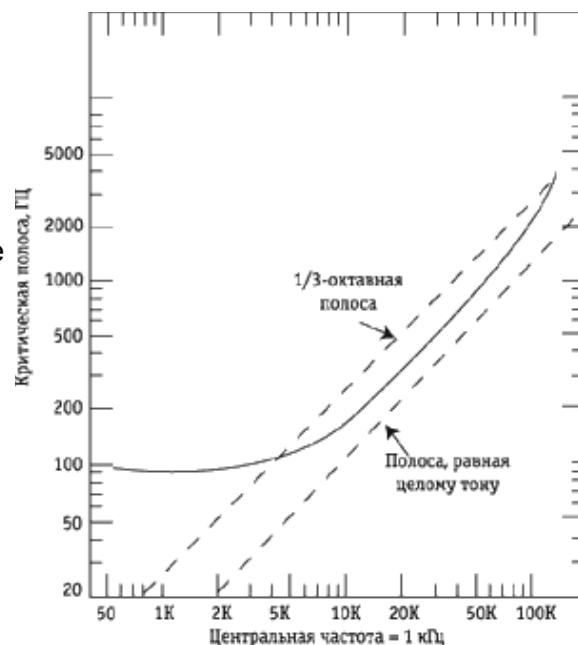


Рис.1 Зависимость ширины третьоктавных полос и критических полос слуха от частоты

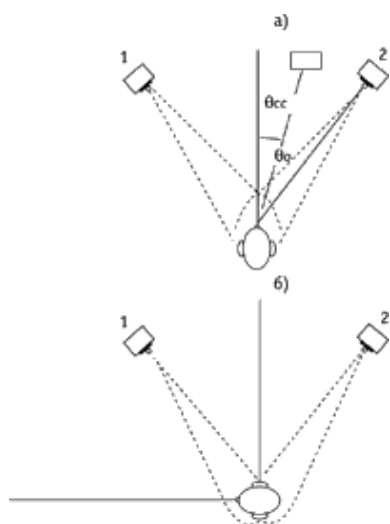


Рис.3 Система расположения громкоговорителей

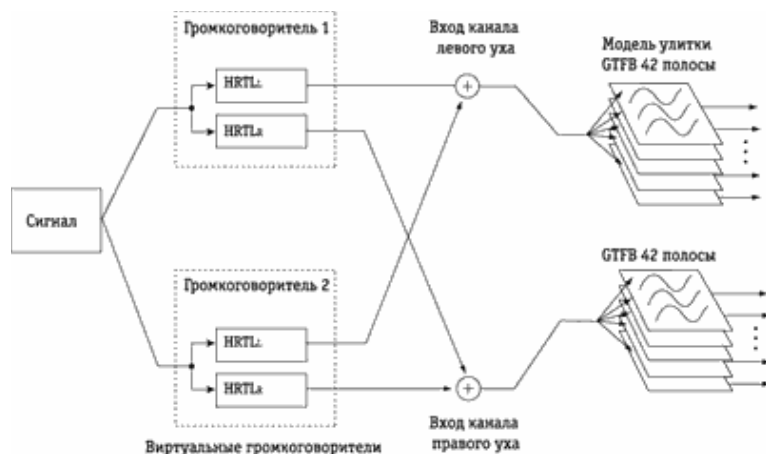


Рис.2 Принцип работы компьютерной слуховой модели

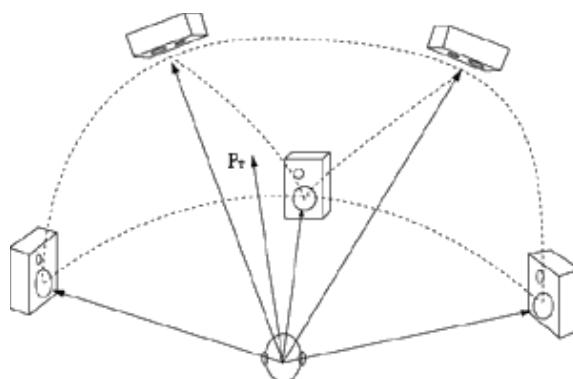


Рис.4 Трехмерная система расположения громкоговорителей

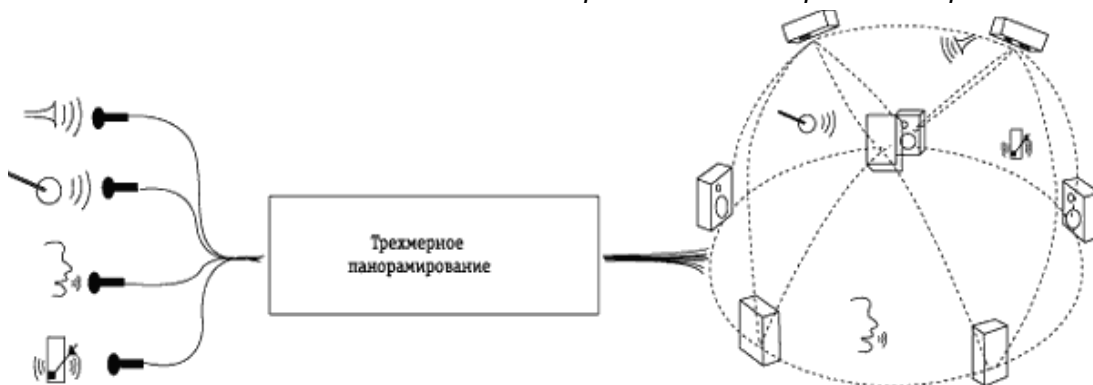


Рис.5 Панорамирование пространственного источника

Кроме того, как было показано в одной из предыдущих статей ("Звукорежиссер", 10/1999), при локализации источников, находящихся в боковой плоскости, существует так называемый "конус неопределенности", где локализация происходит с большим трудом, не подчиняется вышеуказанным законам, носит очень индивидуальный характер, и обычно требует дополнительных поворотов головы.

Для выполнения сравнительной оценки тембров виртуальных и реальных источников на первом этапе определялось положение виртуального источника для заданной конфигурации громкоговорителей расчетным путем по вышеуказанным формулам, а также на компьютерных моделях слуховой системы, с помощью которых производился расчет спектра звукового сигнала в ушном канале в каждой критической полосе (аналогично определению средней громкости). Затем в эту же позицию помещался реальный источник, и для него также проводился аналогичный расчет и дополнительные измерения спектра реального звука в слуховом канале. После этого из спектра мнимого источника вычитался спектр реального источника, и по полученной разности оценивалась разница спектров и, соответственно, разница в воспринимаемых тембрах. Конечно, одной информации о спектрах недостаточно для полного суждения о разнице в тембрах, однако для

первичной оценки это может служить основанием.

Результаты, полученные для двух громкоговорителей в стереопаре в заглушенной камере, представлены на рисунке 6. Здесь показана разность между суммарными сигналами, поступающими в каждое ухо с учетом передаточных функций головы, определяемых дифракцией на голове и ушной раковине. Они были измерены у двадцати различных слушателей от мнимого и реального источника, помещенного на место локализованного мнимого источника. По оси абсцисс отложены центральные частоты для каждой критической полосы, по оси ординат – уровень громкости в фонах (значение громкости в фонах равно значению уровня сигнала в дБ на частоте 1 кГц).

Как видно из рисунка 6, эта разница очень значительна в области 2 кГц (имеется широкий провал более 10 фонов), имеются также существенные различия и в области высоких частот: на частоте 8 кГц – более 4 дБ. Эта разница появляется за счет так называемого эффекта "гребенчатой фильтрации", которая возникает при сложении когерентных (одинаковых) сигналов от двух (или нескольких) громкоговорителей на каждом слуховом канале. Этот эффект происходит при определении положения мнимого источника, и отсутствует, когда звук идет от одного громкоговорителя, находящегося в этом же месте – при таком сложении появляются глубокие провалы в спектре суммарных сигналов.

Аналогичные эксперименты и расчеты были повторены для реальных условий в реверберирующем помещении. Время реверберации было выбрано $0,3 \pm 0,05$ с, а громкоговорители размещались в горизонтальной плоскости под углом 30° на расстоянии 2 м. При расчетах разности в спектрах реального и мнимого источников на модели слуховой системы учитывались бинауральные импульсные характеристики помещения, результаты показаны на рисунке 7. Видно, что наличие в комнате отраженных сигналов уменьшает эффект гребенчатой фильтрации, а также уменьшает разницу между спектрами реальных и мнимых источников – однако она все равно достаточно велика в области средних частот. Если подавить прямой сигнал и первые отражения на промежутке 50 мс, т.е. оставить только диффузный отрезок реверберации, то видно, что никакой разницы (и, соответственно, окрашивания звука, изменения тембра) практически нет, поскольку в диффузном поле эффект гребенчатой фильтрации отсутствует.

Все результаты расчетов и экспериментов были проверены с помощью субъективных экспертиз. На первом этапе прослушивания производились для стереопары, установленной в заглушенной камере под углом $\pm 30^\circ$ на расстоянии трех метров. На место мнимого источника устанавливался сначала реальный источник с определенным уровнем звукового давления, затем слушателей просили установить усиление таким образом, чтобы субъективно выровнять громкости в каждой полосе частот. Испытания производились на разных сигналах: синусоида, узкополосный шум, импульсные сигналы и др.

Результаты показаны на рисунке 8, и они совпадают с результатами, полученными на моделях слуховой системы: в области 2 кГц имеет место подъем усиления примерно на 6 дБ (по-видимому, для компенсации провала за счет эффекта гребенчатой фильтрации); в области низких частот разницы практически нет, поэтому дополнительное усиление для мнимого источника практически оказывается равным нулю (ось ординат на рисунке 8).

Таким образом, слушатели по-разному воспринимают спектральную громкость реального и мнимого источника и, соответственно, по-разному оценивают его тембр.

Испытания были повторены для обычной комнаты прослушивания. Использовались два монитора фирмы Genelec на расстоянии двух метров от слушателей, которые располагались в центре. Результаты тестов для узкополосных шумовых сигналов показаны на рисунке 9. Из полученных результатов видно, что в области 2 кГц мнимый источник имеет громкость ниже, и требуется его усиление – как и в заглушенной камере, но разница меньше. Однако в области низких частот мнимый источник кажется выше по уровню, чем реальный источник в этой же точке (на оси), и поэтому его уровень усиления понижается. Таким образом, помещение снижает отклонения за счет эффекта гребенчатой фильтрации, но вносит свои дополнительные искажения.

Полученные результаты отражают субъективно воспринимаемое различие в спектрах мнимых и реальных источников, а это служит основой для различий в восприятии тембров.

Среди специалистов по звуку существует мнение, что если сравнивать мнимый источник не с реальным источником, находящимся на его месте (например, в центре), а с одним из боковых громкоговорителей из стереопары, то различие в тембре проявляется не так сильно. Действительно, проведенные эксперименты показали, что различие с субъективно воспринимаемой спектральной громкостью будет меньше, если сравнивать мнимый источник в центре с одним из боковых громкоговорителей. Получается интересное психоакустическое явление: слух локализует мнимый источник в одном месте (например, в центре), но формирует представление о тембре, ориентируясь на реальный источник, находящийся сбоку (то есть на громкоговоритель, воспроизводящий этот звук), а не на тембр реального громкоговорителя, который должен был быть в центре.

Испытания были продолжены для панорамирования мнимого источника под разными углами к оси: 0° , 10° , 20° и 30° . Чем ближе мнимый источник подвигается к одному из громкоговорителей (которые расположены под углами $\pm 30^\circ$), тем меньше проявляются отличия в спектрах и, соответственно, окрашивание

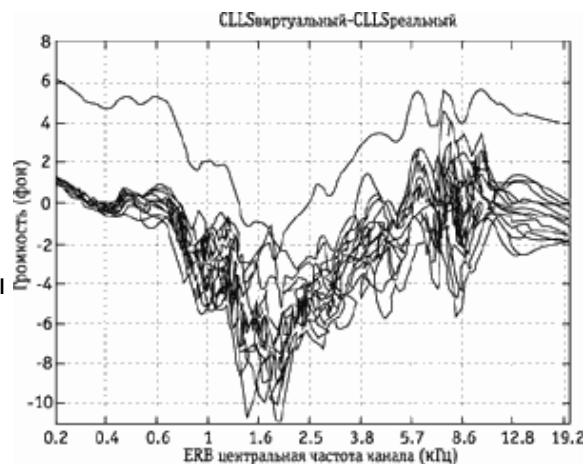


Рис. 6 Разность спектральных уровней громкости реального и мнимого источника

тембра. Кроме того, было исследовано влияние поворотов головы слушателя, находящегося в центре под углами 0°, 30°, 60° к направлению на мнимый источник. Установлено, что область наибольшего окрашивания при поворотах головы смещается в район 4...8 кГц, что соответствует области максимального проявления эффектов дифракции на ушной раковине.

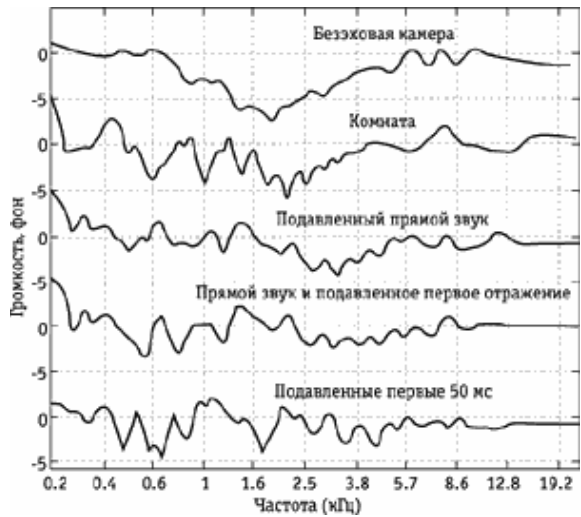


Рис.7 Разность спектральных уровней громкости реального и мнимого источника в заглушенной камере, реальном помещении, при подавлении первых отражений

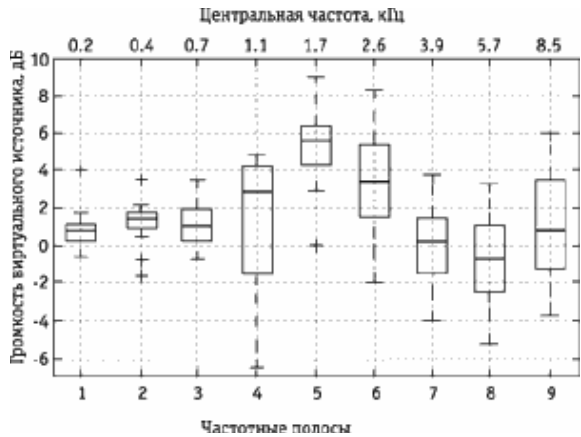


Рис.8 Разность спектральных уровней громкости по результатам субъективных экспертиз в заглушенной камере (8 экспертов)

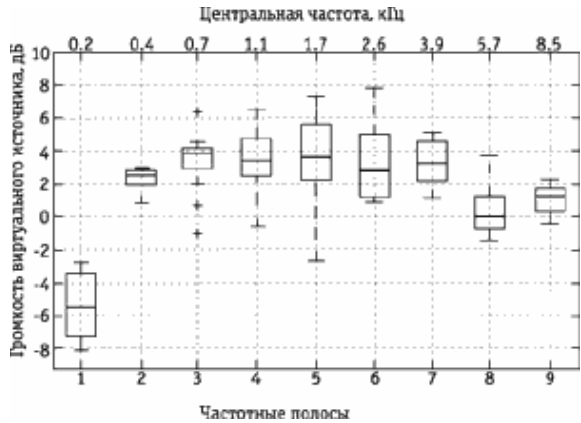


Рис.9 Разность спектральных уровней по результатам субъективных экспертиз в реальном помещении

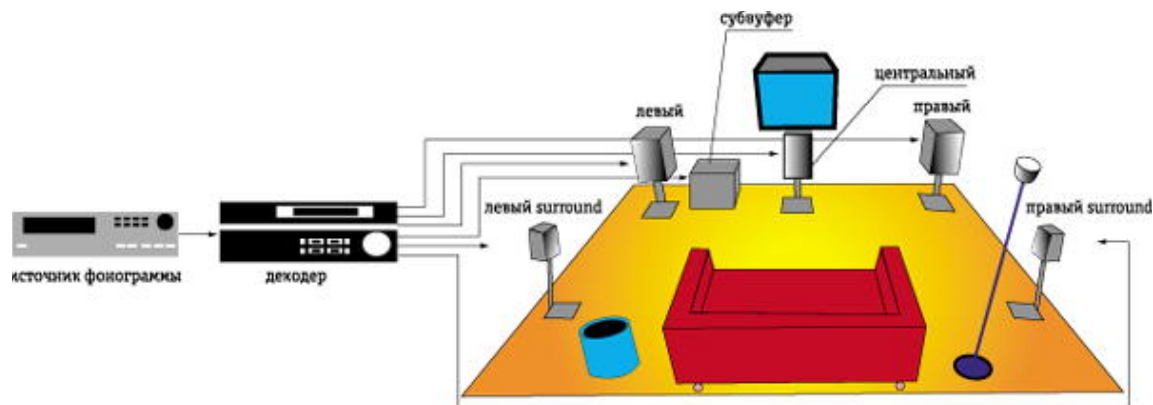


Рис.10 Расположение громкоговорителей по системе Home Theatre

Можно предположить, что когда для создания мнимого источника (распределенного виртуального пространственного образа) используется несколько громкоговорителей, размещенных различным образом в трехмерном пространстве, например, в системах Home Theatre (рисунок 10), то окрашивание, т.е. отличие в воспринимаемом тембре мнимого источника, будет значительно больше. Во-первых, потому, что большее число когерентных сигналов прибывает на каждый ушной канал от нескольких громкоговорителей, поэтому эффекты гребенчатой фильтрации проявляются гораздо сильнее. Во-вторых, когда несколько громкоговорителей разнесены дальше в пространстве, их отличия друг от друга по тембру будут сильнее, и это также увеличит окрашивание мнимого источника. Эти предположения были проверены на моделях слуховой системы, и результаты тестов прослушивания будут опубликованы в ближайшее время.

Громкоговорители были расположены по системе треугольник (рисунок 11) под углами 0° , $+30^\circ$, -30° в горизонтальной плоскости, и $0^\circ +15^\circ$, -15° в вертикальной плоскости. Мнимый источник перемещался в точки, указанные на рисунке 11 кружком: $(0, 0)$, $(0, +15)$, $(0, -15)$. Результаты измерений показаны на рисунке 12. На оси абсцисс отложен уровень громкости в фонах, по оси ординат – центральная частота, как на предыдущем рисунке. На рисунке 11 видно, что провалы за счет гребенчатой фильтрации стали значительно шире, от 1,5 до 4 кГц, при изменении угла подъема они смещаются по частотной шкале. Естественно, что это увеличивает разницу при субъективном прослушивании в уровнях громкости, и требует большей компенсации, что и приводит к большему окрашиванию тембра.

В целом эксперименты по оценке тембров виртуальных источников при размещении различных конфигураций систем излучателей в заглушенной камере и в реальном помещении, выполненные как на моделях слуховой системы, так и с помощью субъективных тестов, показали, что результаты, полученные на моделях и с помощью слушательских тестов, достаточно хорошо коррелируют друг с другом, и могут служить основой для предсказания изменения тембров мнимых источников при различных установках громкоговорителей.

Проведенные эксперименты позволили сделать ряд важных для практики звукозаписи выводов:

- окрашивание тембра виртуального источника при различном положении громкоговорителей происходит за счет двух эффектов: гребенчатой фильтрации и влияния реверберационных процессов в помещении.

Звуковые волны от нескольких громкоговорителей прибывают к ушным каналам в разное время, что служит причиной интерференционных эффектов типа "гребенчатого фильтра" при восприятии мнимого источника. При стереорасположении громкоговорителей этот эффект создает глубокие провалы в области 1...4 кГц в спектре сигналов, поступающих на ушные каналы. Эти провалы приводят к существенным отличиям при восприятии тембра и громкости виртуального и реального источников, особенно при прослушивании в заглушенной камере или сильно заглушенном помещении;

- при увеличении заглушения помещения, то есть при сильном уменьшении времени реверберации, разница в восприятии тембров увеличивается;

- влияние этого эффекта особенно сильно, когда громкоговорители находятся во фронтальной плоскости перед слушателем; когда же они сдвигаются в сторону или слушатель поворачивает голову, то провалы сдвигаются в сторону высоких частот, и становятся менее заметными на слух;

- когда пространственные системы размещаются в реальном помещении, эффект гребенчатой фильтрации проявляется меньше – в диффузном поле он вообще не проявляется, но на восприятие тембра виртуального источника начинают влиять процессы ранних отражений и поздней реверберации в помещении;

- измерения, выполненные для двух громкоговорителей, для системы из трех громкоговорителей ("треугольник"), а также с использованием разных видов сигналов (шумовые, импульсные и др.), показали, что общие закономерности восприятия тембров и громкости виртуального источника сохраняются.

Таким образом, воспринимаемый пространственный виртуальный источник будет по-разному восприниматься в разных помещениях, и степень его различия от реального источника может существенно меняться в зависимости от степени заглушенности помещения. В настоящее время идет активный процесс перехода на системы пространственного многоканального воспроизведения, и соответственно происходит процесс переоснащения контрольных комнат студийными мониторами, расположенными по системе 5.1 (или другими вариантами систем Surround Sound). Поэтому регулирование панорамы, баланса и других параметров

происходит при слуховой оценке именно виртуального пространственного звукового образа. В связи с этим изменения его тембрального окрашивания при изменении положения слушателя и условий в окружающем помещении, а также процесс увеличения отличий его тембра от тембра реального источника, имеют существенное значение для качества звукозаписи. Все это требует продолжения серьезных исследований в данной области, и разработки рекомендаций для практической работы.

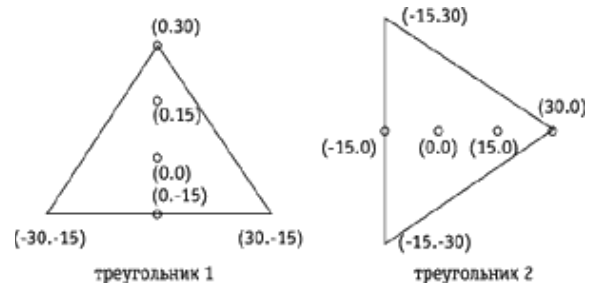


Рис. 11 Расположение громкоговорителей по системе "треугольник"

Часть 16. Взаимодействие акустических систем и помещения - стереофоническое прошлое и многоканальное будущее

Основной темой всех последних конгрессов AES, как уже отмечалось в предыдущих статьях, было рассмотрение различных аспектов внедрения пространственных систем звукозаписи, звукопередачи и звуковоспроизведения.

На предпоследнем 109-м конгрессе выступил с большой лекцией известнейший в аудиомире специалист Флойд Туле (корпорация Harman, США). Лекция называлась "Акустика и психоакустика громкоговорителей и помещения: стерео-прошлое и многоканальное будущее". После этой лекции была развернута большая дискуссия по поднятым в ней проблемам, в ней участвовали всемирно известные специалисты. Поднятые вопросы были настолько интересны, что я решила подготовить обзорную статью по обсуждаемым там проблемам, добавив ряд своих соображений. Именно этим и обусловлено несколько необычное название статьи.

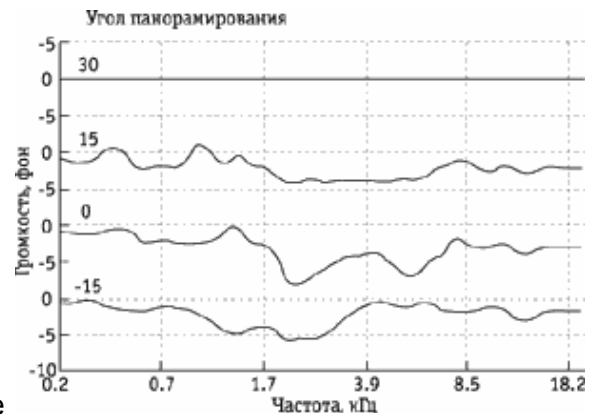


Рис. 12 Расположение громкоговорителей по системе "треугольник"

Первую проблему, которая была поставлена Ф. Туле, и вызвала оживленную дискуссию, можно назвать так – **"Музыка и кино – это искусство, аудио – это наука"**. Тезис, что вся аудиопродукция должна звучать хорошо, кажется очевидным, с ним все согласны. Однако определение того, что такое "звучит хорошо" – дело сложное и противоречивое. Некоторые считают, что это дело персонального вкуса, и оно также разнообразно, как вкус к музыке, вину и пр. Такой подход помещает производителей аудио в категорию людей искусства, которые пытаются обращаться к изменяющимся вкусам публики.

Другие имеют более прагматичный подход (который также разделяю я), считая, что аудиоиндустрия – это только проводник искусства, и ее задача – записывать, сохранять и воспроизводить произведения звукового искусства с той максимальной точностью, с которой это позволяет делать технология. Но не следует забывать, что эта индустрия создала такую новую творческую профессию, как звукорежиссер, который получил, благодаря современным технологиям, большие возможности в моделировании звукового пространства, динамики, тембра и пр.

Для понимания того, что такое "хороший" звук, чрезвычайно важны результаты исследований в психоакустике, с тем, чтобы понять связь между тем, что мы слышим, и тем, что мы измеряем. Инженеры создают аудиопродукцию, опираясь на данные измерений определенных параметров, и задача аудиоиндустрии состоит не в том, чтобы довести эти параметры до предельного уровня (что дорого и неоправданно), а в том, чтобы довести их до слышимого уровня. В связи с этим значительные усилия должны быть направлены на изучение проблемы взаимодействия системы "акустические устройства – помещение – слушатель", с учетом того, что их следует рассматривать как единую звуковую систему со сложными связями внутри нее.

Поскольку проблему расшифровки "звукового образа" в слуховой системе нельзя еще считать решенной (и неизвестно, когда она еще будет решена), то в настоящее время существуют две параллельные системы измерения: объективная и субъективная, которые дополняют друг друга. Именно на их взаимодействии основаны оценки качества звука: как хорошего, или как плохого.

В связи с этим встает другая проблема, которую можно обозначить как **"Слушательские тесты – превращение мнений в факты"**. До настоящего времени широко распространенным было мнение, что субъективные оценки отражают просто разницу вкусов и слуховые способности различных слушателей. Эти факторы, конечно, оказывают влияние, однако при правильно организованных слуховых экспертизах и подборе тренированных слушателей с проверенными слуховыми порогами, их оценки становятся удивительно стабильны, а различие в определении "хорошего" или "плохого" звука очень мало. Добиться этого не так просто, именно поэтому проблемам организации субъективных экспертиз уделяется столько внимания в современных исследованиях. В настоящее время совместными усилиями ряда стран создается новое

поколение стандартов по организации субъективных экспертиз пространственных систем звуковоспроизведения, о которых постараемся рассказать в дальнейшем.

Важнейшими факторами, определяющими стабильность и достоверность субъективных оценок, являются следующие:

- условия прослушивания;
- уровни громкости;
- порядок предъявления материала;
- отбор программ;
- отсутствие визуальных признаков,

а также многое другое, что собственно и определяется в стандартах. Но особенное внимание следует обратить на отбор экспертов – как показали исследования, одним из важнейших факторов, влияющих на субъективную оценку качества звучания, является состояние их аудиометрических данных, т.е. соответствие порогов слуха в широкой области частот стандартным пороговым кривым (рис.1).

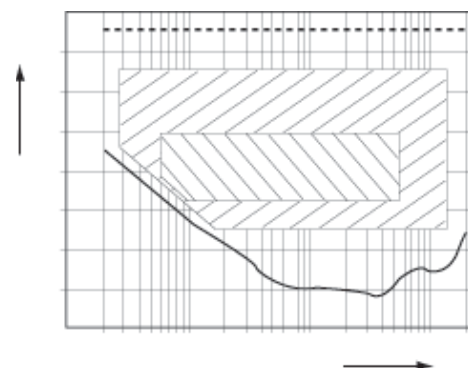


Рис.1 Пороговые кривые слуха

На рисунке 2 показаны результаты субъективных экспертиз для двух групп слушателей: одна с проверенным слухом, соответствующим стандартным порогам, другая – с определенными отклонениями слуха от пороговых значений. Слушателей обеих групп попросили оценить по 10-бальной шкале четыре акустические системы: А, В – очень хорошего качества, С, D – среднего. Для группы с нормальным слухом видна четкая корреляция в

оценках с качеством акустических систем (на рис.2 – это сплошные точки). Для второй группы (рис.2 – контурные точки) при переходе к системам С и D разброс в оценках резко увеличивается, проявляется четкий индивидуализм в простановке баллов. На основе таких оценок невозможно судить о качестве звучания этих систем – слушатели скорее хорошо оценивают те акустические системы, которые лучше компенсируют недостатки их слуха. Поэтому для оценки качества звучания акустических систем можно использовать только тех слушателей, которые имеют отклонения по порогам от нулевого уровня на аудиометрической кривой не больше 20 дБ во всем диапазоне частот, – это будет примерно 75% от всех потребителей.

Эта проблема имеет огромное значение, так как мнение потребителей формируется очень часто на основе мнений людей, занимающихся музыкальной индустрией или "около нее". Особенно это относится к некоторым российским журналам по аудиотехнике, где регулярно обсуждается качество звучания аудиотехники, часто на основе индивидуальных прослушиваний отдельными лицами, которые вообще никогда не проверяли свои слуховые пороги (или заведомо имеют их выше нормы, в силу возрастных изменений или регулярных перегрузок слуха высокими уровнями). Как видно из приведенных выше результатов, мнение таких слушателей отражает только их индивидуальные вкусы и особенности слуха, но не соответствует реальному качеству звучания аппаратуры.

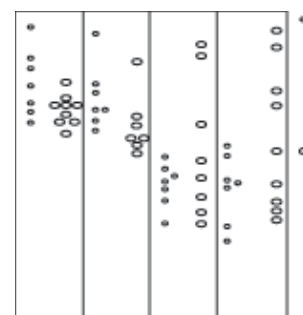


Рис.2 Разброс в оценках качества звука четырех акустических систем различными группами слушателей

Второй важной проблемой является проблема тренировки экспертов: очень часто даже опытные музыканты и звукорежиссеры, участвующие в прослушиваниях, не могут понять, на что именно они должны обращать внимание при прослушивании (многие слышат качество **игры** музыкантов и качество **записи**, а не специфические особенности акустических систем) и не могут выразить свои впечатления словами. Сейчас разрабатывается словарь терминов для субъективной оценки, и созданы компьютерные программы для тренировки экспертов, которые помогают им идентифицировать именно недостатки аудиоаппаратуры.

Многие опытные эксперты (музыканты, звукорежиссеры, инженеры) утверждают, что они могут объективно оценивать качество звучания акустических систем, абстрагируясь от их внешнего вида, размеров и пр. Однако, эксперименты по прослушиванию группы из четырех акустических систем с опытными слушателями (результаты представлены на рис.3) показали, что оценки качества звучания при прослушивании вслепую либо с визуальным контролем сильно отличаются: оценки первых двух систем (дорогих, с хорошим дизайном) сильно повысились, а оценки третьей системы (маленькой) сильно понизились. Таким образом, зрительные впечатления очень активно влияют на слуховую оценку качества, что надо учитывать при желании получить ее максимально объективной.

Известно, какое влияние оказывает помещение на качество звучания акустических систем (полнота, прозрачность, пространственность и т.д.), однако раньше основное внимание оказывалось акустике концертных залов, больших студий и др. В настоящее время, когда общее качество акустических систем значительно улучшилось, влияние небольших помещений прослушивания на субъективные оценки стало предметом серьезных исследований.

Эксперименты в специально спроектированной комнате прослушивания, параметры которой, а также положение акустических систем, подача программного материала и пр., находились под компьютерным контролем, – показали значимость для субъективных оценок положения и способов установки систем, положения слушателей (рис. 4). Результаты экспериментов подтверждают, что только при строгом контроле всех параметров можно добиться повторяемости и статистической значимости субъективных оценок качества звучания, т.е. перейти от "слушательских тестов" к субъективным измерениям.

Следующую обсуждаемую проблему можно назвать так: "Объективные измерения – что мы измеряем, и что мы слышим?". В предыдущие периоды развития аудиотехники существовал огромный разрыв между объективными измерениями и субъективными оценками. Измерялось сравнительно небольшое количество параметров (АЧХ, КНИ, характеристика направленности и др.), измерения проводились недостаточно точно, и результаты измерений не очень хорошо коррелировали с субъективными оценками (по собственному многолетнему опыту разработок акустической аппаратуры в ИРПА могу сказать, что примерно половина времени в процессе разработки акустических систем уходила на обеспечение требований ТУ по объективным параметрам, и еще столько же, если не больше, – на обеспечение хороших субъективных оценок при оценке качества звучания).

Сейчас ситуация значительно изменилась – компьютерные методы измерений позволили значительно увеличить число измеряемых параметров, повысить точность измерений, обеспечить компьютерное моделирование различных условий в помещениях для измерений и т.д. Задача психоакустики на данном этапе – ответить на вопрос: какие именно, и сколько параметров надо измерять в аудиоаппаратуре, с тем, чтобы с их помощью можно было адекватно судить о качестве звучания. Иначе говоря, установить связь между тем, что мы измеряем, и что мы слышим.

Например, наиболее типичный метод объективных оценок акустических систем основан на оценке амплитудно-частотных характеристик (АЧХ), измеренных по оси в заглушенной камере. Даже если эта кривая почти идеально ровная (например, с неравномерностью ± 1 дБ, что уже достигнуто в современных акустических системах), в реальном помещении на прямой звук, измеренный в заглушенной камере, накладываются первые отражения от пола, потолка, стен, а также реверберирующие звука. В результате, вид АЧХ в реальном помещении существенно меняется: повышается уровень на низких частотах, появляются пики и провалы за счет резонансов комнаты и др. Усредненная кривая частотной зависимости акустической мощности от частоты для трех позиций системы, измеренная на четырех различных местах для слушателей, показана на рисунке 5. Отчетливо видно влияние резонансов помещения в области ниже 500 Гц – таким образом, в области низких частот помещение оказывает особенно существенное влияние на субъективные оценки, и, следовательно, измерений АЧХ на оси в заглушенной камере явно недостаточно. Для предсказания того, что мы слышим, надо оценивать АЧХ в помещении на различных слушательских местах, а также измерять частотное и пространственное распределение акустической энергии в помещении. Это и позволяет сделать внедряемое сейчас в практику измерений "распределение Вигнера". Нужно отметить, что в профессиональных целях (кинозалах, театрально-концертных системах звукоусиления и др.) акустические системы давно оценивались с помощью измерений АЧХ в различных точках помещения на шумовом сигнале. В бытовой технике эти методы только начинают использоваться.

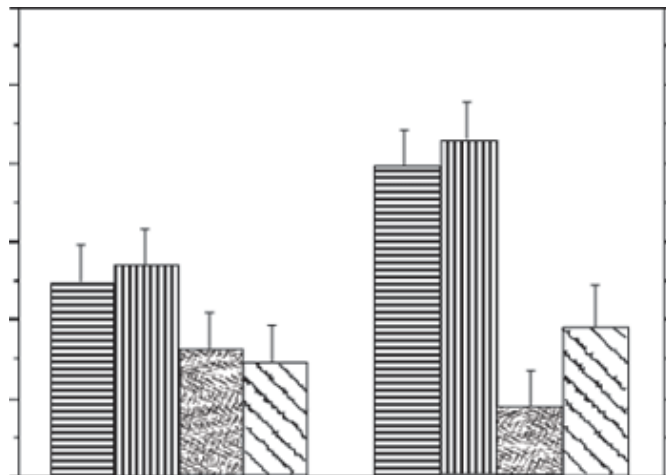


Рис.3 Результаты слуховых оценок при визуальном контроле (слева) и без него (справа)



Рис.3 Вид комнаты прослушивания

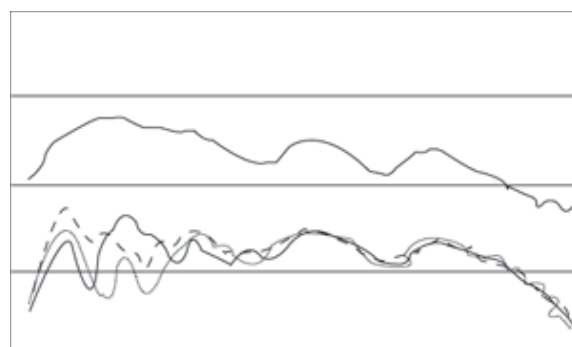


Рис.5 Усредненная частотная характеристика в помещении

В связи с этим встает вопрос о применении эквалайзеров для выравнивая АЧХ в помещении. Этот вопрос давно обсуждается, и тут имеются противоречивые мнения: одни специалисты считают, что это единственный путь решения проблемы "АС плюс помещение", другие вообще избегают их применения. Трудности заключаются в том, что выравнивая АЧХ на оси, можно "испортить" АЧХ под углами, и, соответственно, ухудшить воспринимаемое качество звука. Поэтому применять эквализацию можно, контролируя характер зависимости изменения характеристики направленности от изменения частоты для данной системы.

Далее, выравнивая АЧХ, можно ухудшить фазовые соотношения и, следовательно, временную структуру сигнала, что также может иметь нежелательные последствия. Наконец, выравнивание на низких частотах, где сильнее всего сказываются резонансы помещения, – это вообще трудная задача. Сейчас появились адаптивные цифровые процессоры, которые могут корректировать временную структуру сигнала (амплитуду и фазу соответственно), – они находятся в стадии активного освоения в аудиотехнике и можно надеяться, что с их помощью удастся решить некоторые проблемы.

Организуя объективные измерения, необходимо решить несколько задач: понять акустические и механические процессы, которые необходимо измерять, т. е. понять "что надо измерить"; установить, какие изменения параметров являются слышимыми (это психоакустическая задача); создать измерительные системы, которые могут четко выявлять слышимые признаки. Например, запись АЧХ несет в себе информацию, по крайней мере, о трех видах значимых для субъективной оценки параметров: о спектральном балансе, резонансах, акустической интерференции (рисунок 6). Важность спектрального баланса между низко-, средне- и высокочастотной частями диапазона совершенно очевидна для оценки качества, его контроль по форме АЧХ может быть выполнен даже визуально. Наличие резонансов является чрезвычайно важным для слуховых оценок, потому что наша слуховая система очень чувствительна к резонансам (их местоположению и форме), поскольку распознавание фонем (звуков) речи происходит именно по положению резонансов (формант), также как распознавание тембров различных музыкальных инструментов в значительной степени зависит от положения их резонансов. Поэтому наличие резонансных пиков на АЧХ несет очень много информации для слуха.

Наконец, на АЧХ имеется ряд пиков, обусловленных интерференционными явлениями, которые имеют существенно другую значимость для слуха, поскольку, в отличие от реальных резонансов, которые отличаются стабильностью, и слуховая система к ним менее чувствительна (своего рода эффект их перцептуального подавления) – эти смещаются или исчезают при изменении позиции микрофона. Слышимость резонансных пиков на АЧХ в зависимости от их амплитуды, частоты и добротности (ширины) многократно исследовалась на разных видах музыкальных программ (особенно четко слух выявляет наличие пиков на АЧХ на розовом шуме), что позволило установить пороги слышимости для этих видов искажений.

Эксперимент был организован таким образом, что на определенном программном материале, например, звучании симфонического оркестра, в акустические системы подмешивались резонансные пики разной высоты и ширины в различных частотных областях. При этом менялись их параметры и определялись слуховые пороги их заметности. Результаты показаны на рисунке 7, за нулевой принят средний уровень программного материала. Оказалось, что слух наиболее чувствителен к широким пикам (с малой добротностью $Q = 1$) в области 2...3 кГц. Даже, когда амплитуда этих пиков на 20 дБ ниже среднего уровня, они уже замечаются слухом как тембральное окрашивание. Наименее чувствителен слух к узким пикам ($Q = 50$), они могут быть даже на 2...3 дБ выше среднего уровня – отсюда допустимая неравномерность для систем категории Hi-Fi составляет 2 дБ.

В многополосных системах иногда возникает ситуация, когда за счет фильтрации уровень АЧХ, например, низкочастотного громкоговорителя, понижается после частоты среза на 12 дБ/окт, но, если низкочастотный громкоговоритель имел резонансные пики на АЧХ, они могут быть ослаблены до величины меньше, чем 20 дБ,

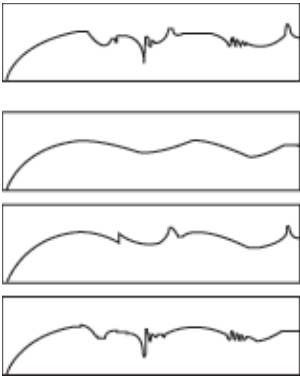


Рис.6 Виды информации в АЧХ

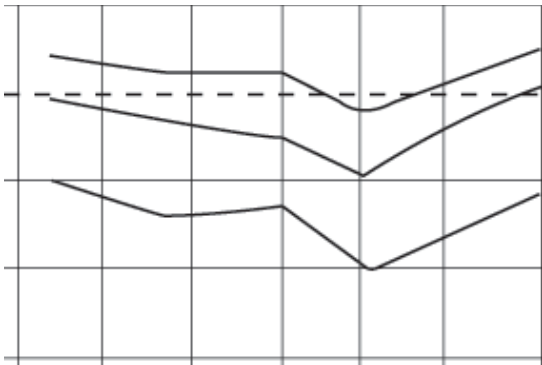


Рис.7 Пороги заметности резонансных пиков

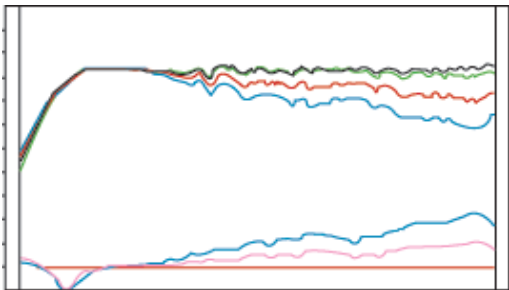


Рис.8 АЧХ и индексы направленности для монитора

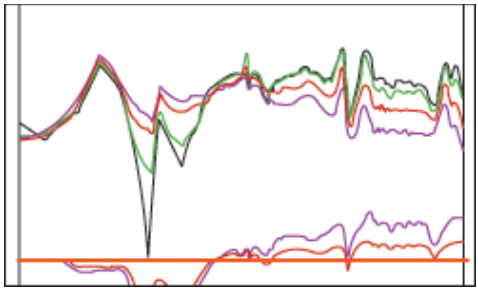


Рис.9 АЧХ и индексы направленности для бытовой АС

и будут восприниматься как окрашивание (чаще всего это замечается на голосе). При этом суммарная АЧХ акустической системы будет гладкой. Отсюда вытекают требования к методикам измерений, обеспечивающим возможность измерять резонансные пики, соответствующие порогам слышимости (для этого необходимо использовать FFT с очень высокой разрешающей способностью).

Это также касается и других параметров, например, гармонических и интермодуляционных искажений. Экспериментально установлены пороги

слышимости гармонических нелинейных искажений на разном программном материале, например, на фортепианной музыке – 1...2%, при этом чувствительность слуха к искажениям нечетных порядков существенно выше, чем к четным. Современная метрология позволяет измерять пороговые значения нелинейных искажений, так что с этим параметром особых проблем нет, но зато мало данных по исследованию "нелинейной компрессии" в акустических системах, т.е. изменения формы АЧХ при изменении уровня подводимой мощности. Это особенно нежелательно для студийных мониторов – если у них будет изменяться форма АЧХ и соответственно тембр при разных уровнях громкости. Установление субъективных порогов по этим видам нелинейности еще требует исследований.

Таким образом, первоочередная задача состоит в том, чтобы определить достаточный перечень измеряемых объективных параметров, и установить их связь с субъективными оценками аппаратуры. На сегодня эта проблема находится на достаточно высоком уровне ее решения: современные цифровые метрологические станции позволяют производить измерения достаточно большого количества параметров (более 30) с очень высокой точностью как в заглушенных, так и незаглушенных помещениях, причем имеется достаточно четкая корреляция этих параметров с результатами субъективных тестов. На рисунке 8 показана АЧХ на оси и под различными углами, а также частотная зависимость индекса направленности I1 (отношение давления прямого звука к общей энергии) и индекса I2 (прямого звука к энергии ранних отражений) для профессионального студийного монитора с очень высокими субъективными оценками, а на рисунке 9 – те же параметры для бытовой акустической системы, получившей средние оценки.

Однако до окончательного решения проблемы, которое позволило бы по измерениям определенного комплекса объективных параметров предсказать качество звучания аудиоаппаратуры, еще много работы для психоакустики.

Наконец, последнюю обсуждаемую проблему можно было бы сформулировать так: **"Акустические системы и помещение вместе – очень сложный комплекс"**.

Во взаимодействии акустических систем и помещения можно выделить две категории характеристик.

Первая – отраженные звуки, которые определяются взаимодействием характеристик направленности излучателей и отражающей способностью помещения (отделкой стен, формой, размерами помещения и т.д.);

Вторая – резонансы помещения, степень возбуждения и восприятия которых зависит от размеров и формы помещения, расположения излучателей и слушателей и др.

Раньше производитель не имел контроля над тем, в каком помещении будет использоваться его аппаратура, а поскольку помещение – полноправный член единой акустической системы, то сейчас ситуация в аудиотехнике начинает меняться:

во-первых, разрабатываются акустические системы с учетом специфики их установки в помещениях,

во-вторых, в распоряжении производителя и потребителя появился новый класс аппаратуры – цифровые адаптивные процессоры, позволяющие улучшить взаимодействие систем и помещения,

в третьих, все больше проявляется тенденция, когда производитель берет на себя оптимальную установку сложных акустических комплексов (например, аппаратуру для домашнего театра) в помещениях у потребителя. Можно ожидать в будущем, что дома уже будут строиться с полным встроенным комплексом аудио-видеоаппаратуры.

Первая задача – взаимодействие характеристик направленности и отраженных звуков от стен помещения, – решалась в разные периоды различными способами. Даже в монофоническую эпоху было много попыток

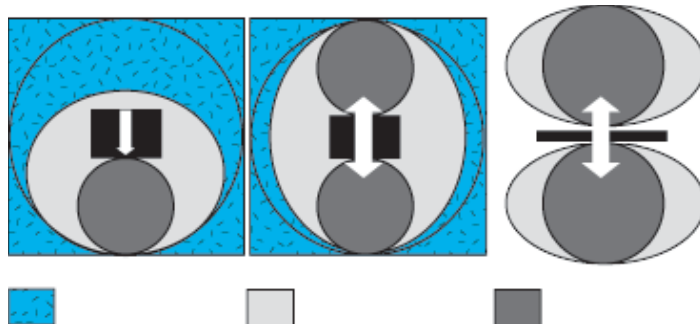


Рис.10 Различные виды характеристик направленности

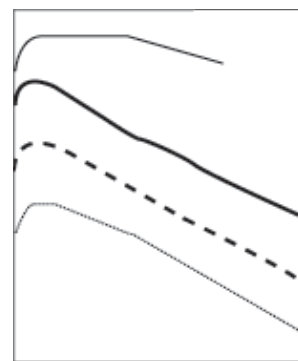


Рис.11 Слышимость отраженных звуков

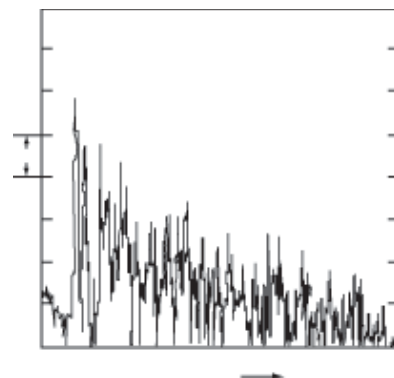


Рис.12 Реверберационная кривая

улучшить это взаимодействие за счет размещения излучателей в углах помещения, создания разных систем псевдостереофонии и т.д. В эпоху стерео появлялись многочисленные попытки изменить соотношения прямого и отраженного звука с целью улучшения пространственных иллюзий у слушателей (создать у слушателя впечатление нахождения в концертном зале). Ближе всего к решению этой проблемы подошла только бинауральная стереофония.

Среди всех используемых в стереофонии излучателей наиболее типичные характеристики направленности представлены на рисунке 10, где показана направленность излучения для низких, средних и высоких частот. Монополь (10а) – обычный излучатель прямого излучения, 10б – биполь, 10в – диполь (например, электростатический излучатель). Установка излучателей с такими разными характеристиками направленности в одном и том же помещении создает у слушателей совершенно разные ощущения пространственности, ясности и полноты, поэтому выбор и согласование характеристик направленности акустических систем и параметров помещения – это большое искусство. Например, в большинстве современных студий, где обычно применяются мониторы прямого излучения, используются сильно заглушенные контрольные комнаты, что позволяет звукорежиссеру точнее контролировать звук. В жилых помещениях необходим компромисс между уровнем прямого и отраженных звуков, который требует большого опыта и знаний в выборе и расстановке излучателей и обработке помещений.

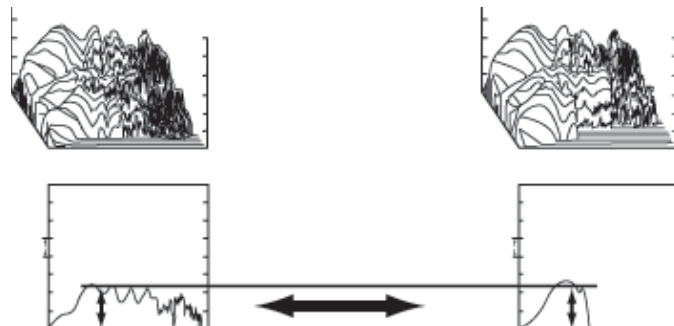


Рис.13 Трехмерные спектры отраженных сигналов

В настоящий период, период внедрения многоканального звука, ситуация меняется – впервые в истории аудиотехники могут быть выбраны такие конфигурация и расположение излучателей в домах потребителей, которые обеспечат им пространственное впечатление, близкое к тому, что было при записи. Сейчас идут дебаты относительно выбора числа каналов (5.1; 7.1, даже 10.2), способов кодирования, расположения излучателей и т.д.

В профессиональной аудиотехнике (в кинопрокате, например) есть стандарты для выбора конфигурации и характеристик направленности излучателей. Для домашнего применения многоканальных систем стандарты сейчас разрабатываются. В отличие от стереосистем, где относительное ощущение пространственного звукового образа создается только в узкой зоне стереоэффекта, от многоканальных систем можно ожидать, что они создадут значительно более широкую зону пространственного звука, окружающую слушателя со всех сторон, менее чувствительную к условиям вторичного помещения.

Для создания таких пространственных иллюзий в малых помещениях необходимо, кроме размещения специальных тыловых (surround) излучателей, учесть то, что добавление электронной декорреляции между левыми и правыми каналами усиливает ощущение пространственности. Учитывая возможности, например, пятиканальных систем, появляется много вариантов для художественного использования этого метода. Возможность размещения слушателей, в основном, в условиях прямого излучения от пятиканальной системы, делает необходимым выбор излучателей со специальными характеристиками направленности, или размещение звукопоглощающего материала на стенах помещения в направлении максимального излучения.

В любом случае необходима разработка новых измерительных методов для оценки допустимого уровня отражений и их влияния на восприятие пространства и других характеристик при использовании многоканальных звуковых систем.

Достаточно много результатов по слышимости отраженных звуков было получено для больших концертных залов и студий, однако для маленьких помещений результатов еще недостаточно. Некоторые результаты по слышимости отраженных сигналов на речи представлены на рисунке 11. Нижняя кривая представляет порог обнаружения отраженных звуков при разных временах задержки (от 0 до 60 мс) и разных уровнях относительно прямого звука (уровень прямого звука принят за 0 дБ). На пороге обнаружения слушатели описывали свои ощущения как "появление пространственности" при сохранении локализации источника на прямой звук. При повышении уровня отраженных звуков на 10 дБ (следующая кривая) слушатели отмечали изменение размеров пространственного образа или его сдвиг, но локализация четко шла по прямому звуку.

Дальнейшее повышение уровня отраженных звуков приводит к ощущению появления второго источника, и затем к смене локализации на отраженный звук. На второй кривой точками обозначены типичные для



маленькой комнаты первые шесть отражений – очевидно, что они попадают на кривую, где слушатели четко отмечают изменение пространственного образа в размерах и его сдвиг, т.е. маленькие комнаты существенно влияют на формирование пространственного образа. Эти кривые получены для речевых и других видов сигналов (музыка, шум и др.). Форма кривых изменяется, но неизменным остается общий вывод – все отраженные звуки влияют на восприятие пространственного образа, т. е. помещение – это полноправный участник звуковой системы.

Отсюда встает вопрос: что нам следует измерять в помещениях прослушивания? Обычно измеряются кривые спада энергии во времени (рисунок 12) – реверберационные кривые. Но такие кривые не дают прямой корреляции со слуховыми ощущениями. Гораздо более четкая связь существует между спектральным уровнем отраженных сигналов в разные моменты времени и их слуховым восприятием. На рис.13 показаны два вида кумулятивных спектров отраженных звуков в определенной точке помещения, которые по-разному воспринимаются на слух, чего нельзя установить по реверберационной кривой. Полезнее снимать эту кривую в разных частотных диапазонах, выбирая испытательный сигнал, отфильтрованный соответствующим полосовым фильтром.

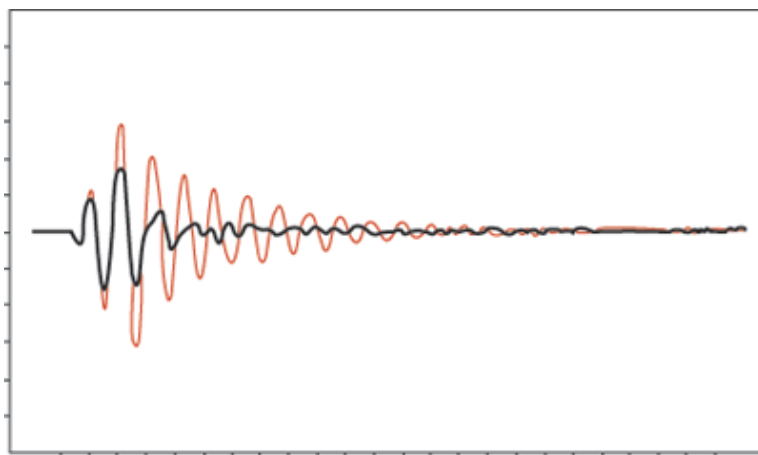


Рис.14 Результаты выравнивания АЧХ и переходных процессов

Второй вид характеристик, о которых было сказано выше, – резонансы помещения. Все помещения имеют резонансы. При этом, чем меньше объем помещения, тем в более высокочастотной части они находятся, например, для помещения объемом 7 x 10 x 5 м первый резонанс помещения равен 17,2 Гц, а для помещения с размерами 3,5 x 5 x 2,5 м первая резонансная частота равна 34,4 Гц, вторая – 68,8 Гц и т.д. В современных многоканальных системах обычно используются низкочастотные блоки (subwoofer), диапазон которых совпадает с областью резонансов помещения, что может существенно изменить тембр звучания низких частот. Поэтому встает серьезная проблема эквализации на низких частотах.

Подходить к этой проблеме нужно с большой осторожностью: во-первых, надо применять измерительные методы с большой разрешающей способностью (не более 0,1 октавы), чтобы уловить эти резонансы, а, во-вторых, использовать точные параметрические фильтры для их вырезания (обычные третьоктавные эквалайзеры не подходят). Поскольку помещение ведет себя как минимально-фазовая система, выравнивание резонансных пиков в амплитудной области приведет к соответствующему уменьшению переходных процессов во временной области. При этом не надо стараться убирать провалы на АЧХ – они обычно вызваны интерференционными эффектами, к ним слух менее чувствителен, и убрать их можно простым сдвигом позиции слушателей. На рисунках 14а,б показан результат эквализации выраженного резонанса на частоте 46 Гц, выявленного при измерениях с высокой разрешающей способностью. Видно, насколько улучшились при этом переходные искажения, соответственно, звук на низких частотах стал более точный, сухой, не бубнящий.

Подводя итоги, можно сказать, что обсуждаемые здесь вопросы, конечно, не решают всей проблемы перехода на многоканальные системы и их взаимодействия с помещением. Однако, в целом, можно заключить, что в понимании этих проблем уже имеется значительный прогресс, и, если он будет продолжаться дальше такими же темпами, можно ожидать, что аудиоиндустрия сможет обеспечить слушателей реальным пространственным звуком в любых помещениях.

Центр внимания сейчас сосредоточен на исследованиях в психоакустике – именно от них зависит темп развития аудиоиндустрии. Поэтому во всем мире психоакустике уделяется такое внимание, и можно только пожалеть, что у нас, при таком интеллектуальном потенциале, вообще нет научных центров, которые бы занимались этой тематикой комплексно.

Часть 17 Слух и речь. Часть 1

От автора

Начиная с этого номера журнала вниманию читателей предлагается серия статей, посвященных слуховому восприятию речевых сигналов (включая вокальную речь, т.е. пение). С одной стороны, запись и обработка речи и пения является одной из труднейших задач в звукорежиссуре, с другой стороны, за последние годы в психоакустике появилось много новых результатов по слуховой оценке и восприятию речи (особенно за прошедшее десятилетие, в связи с задачами обучения компьютера пониманию и синтезу речевых сигналов). Поэтому данная информация может оказаться полезной для практической работы.

Название этой серии статей – "Слух и речь" – появилось не случайно. В 1973 г. под таким же названием было издано хорошее учебное пособие проф. Я.Ш. Вахитова (ЛИКИ). По нему училось не одно поколение студентов, но за прошедшие тридцать лет многое изменилось в понимании процессов создания и восприятия речи.

В представленной серии предполагается последовательно рассмотреть вопросы механизмов образования речи, интегральные и статистические характеристики речевых сигналов, методы их оценки и распознавания, и механизмы восприятия речи (разумеется достаточно кратко и на популярном уровне, поскольку проблемы эти чрезвычайно сложны и еще многое в них остается неразгаданным). В России имеется хорошая научная школа, занимающаяся акустикой и восприятием речи, поэтому издано достаточно много русскоязычных книг: труды Л.А. Чистович, Г.А. Вартамян, В.П. Морозова, В.И. Галунова и др. Имеется также огромное количество монографий и учебников на английском, немецком и др. языках. Для тех, кто глубоко заинтересуется этими проблемами, в конце серии статей будет приведена основная библиография.

Основные механизмы звукообразования речи

Речевой сигнал является средством передачи разнообразной информации как вербальной (словесной), так и невербальной (эмоциональной). Для быстрой передачи информации в процессе эволюции был отобран особым образом закодированный и структурированный акустический сигнал. Для создания такого специализированного акустического сигнала используется "голосовой аппарат", совмещенный с физиологическим аппаратом, предназначенным для дыхания и жевания (поскольку речь возникла на поздних стадиях эволюции, то к речеобразованию пришлось приспособить уже имеющиеся органы

Процесс образования и восприятия речевых сигналов, схематически показанный на рисунке 1, включает в себя следующие основные этапы: формулировка сообщения, кодирование в языковые элементы, нейромускульные действия, движения элементов голосового тракта, излучение акустического сигнала, спектральный анализ и выделение акустических признаков в периферической слуховой системе, передача выделенных признаков по нейронным сетям, распознавание языкового кода (лингвистический анализ), понимание смысла сообщения.



Рис. 1
Основные процессы образования и восприятия речи

Голосовой аппарат является, по существу, духовым музыкальным инструментом. Однако среди всех музыкальных инструментов он не имеет себе равных по своей многогранности, разносторонности, возможности передачи малейших оттенков и др. Все способы звукоизвлечения, которые используются в духовых инструментах, используются и в процессе образования речи (в т.ч. вокальной речи), однако все они перестраиваемы (по приказам мозга), и имеют широчайшие возможности, недоступные ни одному инструменту.

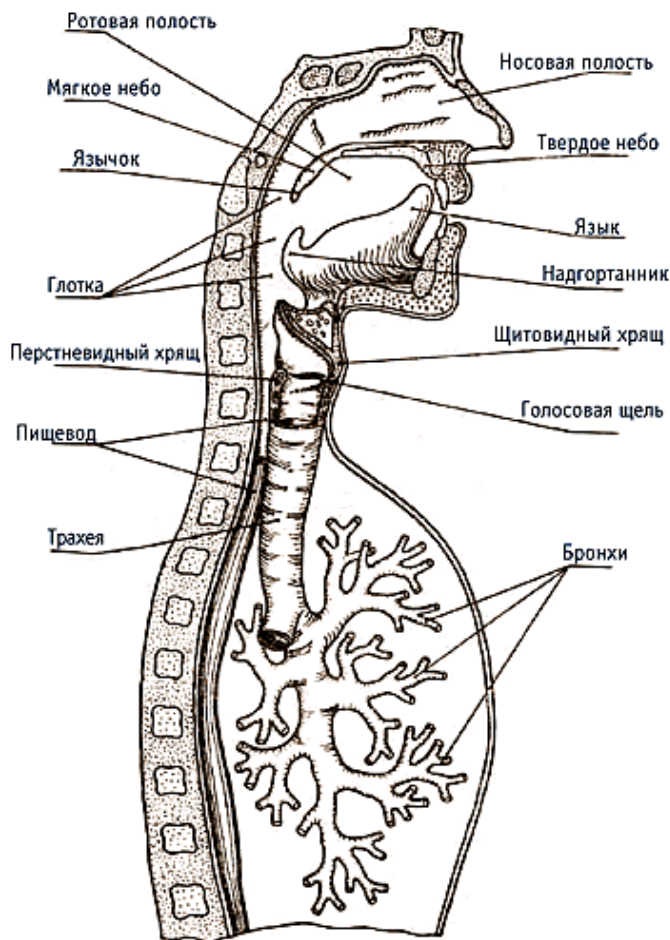


Рис. 2
Структура голосообразующего аппарата

Если рассматривать структуру голосообразующего аппарата как духового музыкального инструмента, он состоит из трех основных частей (рисунок 2):

- **генератора** – дыхательной системы, состоящей из воздушного резервуара (легких), где запасается энергия избыточного давления, мускульной системы и выводного канала (трахеи) со специальным аппаратом (гортанью), где воздушная струя прерывается и модулируется;
- **вибраторов** – голосовых связок, воздушных турбулентных струй (создающих краевые тоны), импульсных источников (взрывов);
- **резонаторов** – разветвленной и перестраиваемой системы резонансных полостей сложной геометрической формы (глотки, ротовой и носовой полости), называемой артикуляционной системой.

Генерация энергии воздушного столба происходит в легких, которые представляют собой своеобразные меха, создающие поток воздуха при вдохе и выдохе за счет разницы атмосферного и внутрилегочного давления. Процесс вдоха и выдоха происходит за счет сжатия и расширения грудной клетки, которые осуществляются обычно с помощью двух групп мышц: межреберных и диафрагмы, при глубоком усиленном дыхании (например, при пении) сокращаются также мышцы брюшного пресса, груди и шеи. При вдохе диафрагма уплощается и опускается вниз, сокращение наружных межреберных мышц поднимает ребра и отводит их в стороны, а грудину – вперед. Увеличение грудной клетки растягивает легкие, что приводит к падению внутрилегочного давления по отношению к атмосферному, и в этот "вакуум" устремляется воздух. При выдохе мускулы расслабляются, грудная клетка за счет своей тяжести возвращается в исходное состояние, диафрагма поднимается, объем легких уменьшается, внутрилегочное давление растет, воздух устремляется в обратном направлении. Таким образом, вдох – процесс активный, требующий затраты энергии, выдох – процесс пассивный. При обычном дыхании этот процесс происходит примерно 17 раз в минуту, управление этим процессом как при обычном дыхании, так и при речи, происходит бессознательно, но при пении процесс постановки дыхания происходит сознательно и требует длительного обучения.

Количество энергии, которое может быть израсходовано на создание речевых акустических сигналов, зависит от объема запасаемого воздуха и соответственно от величины дополнительного давления в легких. Учитывая, что максимальный уровень звукового давления, который может развивать певец (имеется в виду оперный), составляет 100...112 дБ, то очевидно, что голосовой аппарат является не очень эффективным преобразователем акустической энергии. Его КПД составляет порядка 0,2%, как и у большинства духовых инструментов.

Модуляция воздушного потока (за счет вибраций голосовых связок) и создание подглоточного избыточного

давления происходит в гортани. Гортань (larynx) – это клапан, (рисунок 3), который находится на конце трахеи (узкой трубки, по которой воздух поднимается из легких). Этот клапан предназначен для предохранения трахеи от попадания посторонних предметов и для поддержания высокого давления при подъеме тяжестей. Именно этот аппарат и используется в качестве голосового источника при речи и пении. Гортань образована из набора хрящей и мышц. Спереди ее охватывает щитовидный хрящ (thyroid), сзади – перстневидный хрящ (cricoid), сзади также располагаются более мелкие парные хрящи: черпаловидные, рожковидные и клиновидные. Сверху гортани расположен еще один хрящ-надгортанник (epiglottis), также типа клапана, который опускается при глотании и закрывает гортань. Все эти хрящи соединены мышцами, от подвижности которых зависит скорость поворота хрящей. С возрастом подвижность мышц уменьшается, хрящи также становятся менее эластичными, поэтому возможности виртуозного владения голосом при пении также уменьшаются.

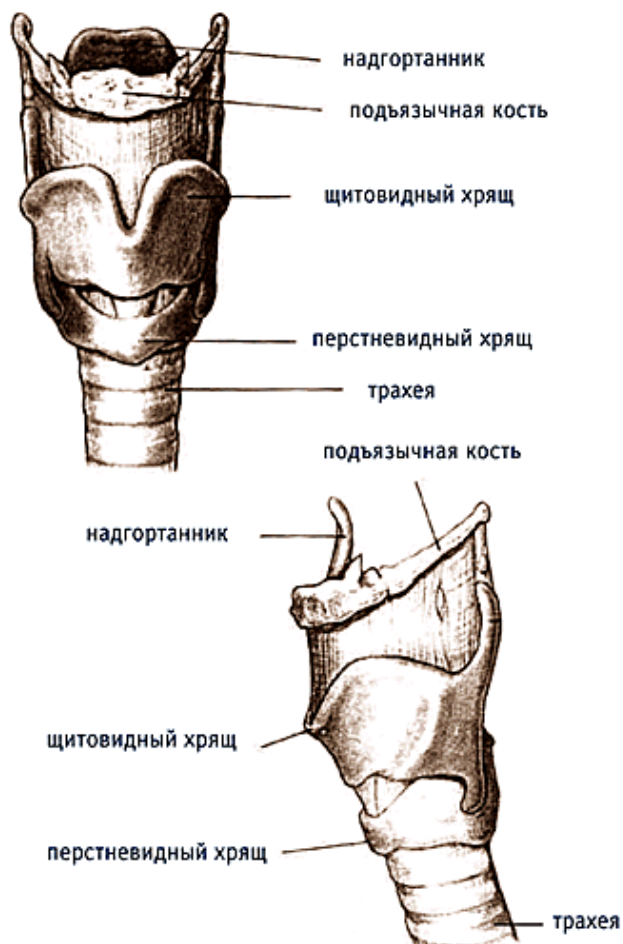


Рис. 3
Строение гортани

Наиболее сложно устроен средний отдел гортани (рисунок 4), в котором расположены парная мышечная перегородка (эластичный конус) и две пары складок. Верхние называются преддверными, или "ложными голосовыми", а нижние – голосовыми. В толще последних лежат голосовые связки, образованные эластическими волокнами, и мышцы (рисунок 5). Промежуток между правой и левой голосовыми складками называется голосовой щелью. Голосовые связки натянуты между щитовидным и черпаловидным хрящами. Размеры голосовой щели в открытом состоянии 2 см в длину и 1 см в ширину.

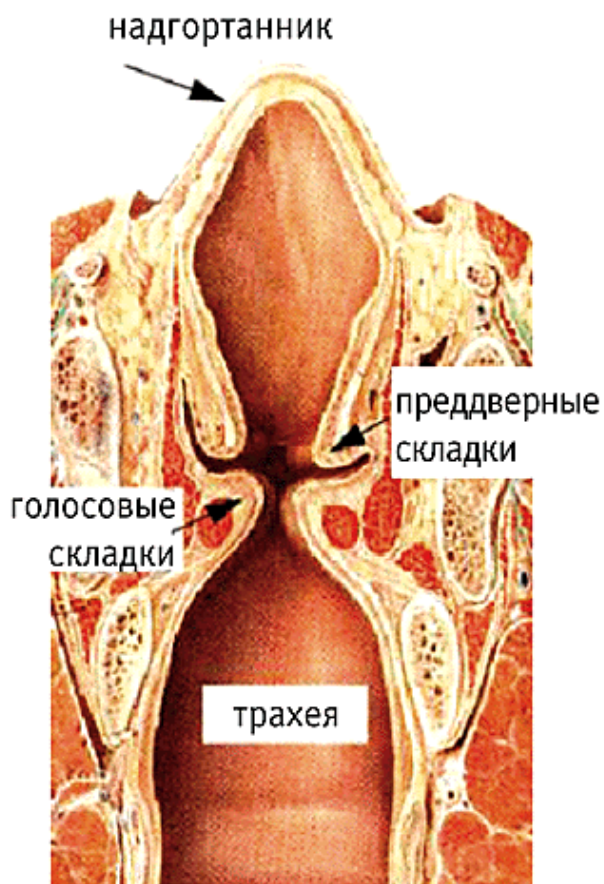


Рис. 4
Сечение трахеи и гортани

Именно **голосовые складки** и являются основным (но не единственным) источником голосообразования (вибратором). Преддверные голосовые складки выделяют специальную слизистую секрецию, которая помогает смазывать голосовые складки и предохраняет их от повреждения при трении во время звукообразования. Обычно они не участвуют в процессе звукообразования, однако при некоторых патологиях истинных связок, они могут участвовать в образовании звука (например, пение Луи Армстронга). (*Хрипота голоса Армстронга была вызвана бородавчатыми образованиями на голосовых связках – это лейкоплакия, проявляющаяся как участки ороговения эпителия. Диагноз "лейкоплакия" был поставлен артисту в зрелом возрасте, но хрипота в голосе присутствует уже на его первых записях, сделанных в возрасте 25 лет. – прим. ред.*).

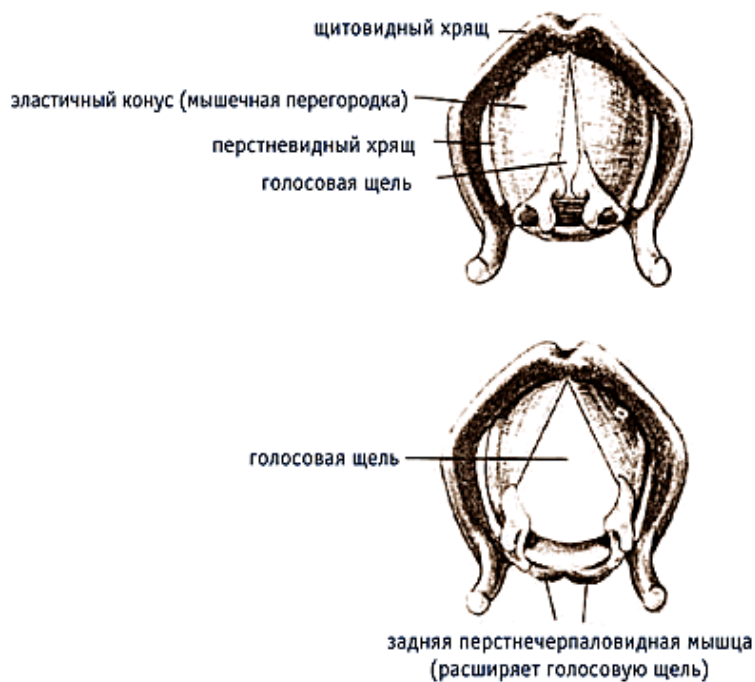


Рис. 5
Голосовая щель и голосовые складки в динамике

Между двумя парами складок находятся небольшие полости (желудочки гортани), которые позволяют беспрепятственно голосовым складкам и играют роль *акустических фильтров*, уменьшая уровень высоких

гармоник (скрипучесть голоса), они же играют роль резонаторов для тихих тонов и при пении в фальцете. При движении черпаловидных хрящей голосовые складки могут сдвигаться и раздвигаться, открывая проход воздуха. При поворотах щитовидного и перстневидного хрящей они могут растягиваться и сжиматься, при активации вокальных мышц они могут расслабляться и напрягаться. Процесс образования звуков речи определяется движением (колебаниями) связок, что приводит к модуляции потока воздуха выдыхаемого из легких. Такой процесс называется *фонацией* (существуют и другие механизмы звукообразования, они будут рассмотрены дальше).

Начнем с рассмотрения **процесса фонации**: перед началом речи голосовые складки должны быть сведены черпаловидными хрящами, что приводит к запираанию потока воздуха и возникновению избыточного подглоточного давления (происходит "предфонационная настройка"). Воздух, который выталкивается легкими из трахеи, накапливается в подскладочном пространстве, и начинает давить на них. Когда избыточное давление повышается до определенной величины, складки размыкаются и воздух устремляется в голосовую щель. В момент максимального открытия щели скорость потока воздуха становится максимальной, давление внутри щели падает (по закону Бернулли), причем скорость протекания воздуха неодинакова – в самой узкой части голосовой щели она максимальна. Внутри голосовой щели образуется зона пониженного давления. Окружающее более высокое давление, а также собственная упругость связок заставляют складки сомкнуться. Этот процесс аналогичен возбуждению колебания тростей в деревянных духовых инструментах. Таким образом, чередование избыточного давления в подскладочном пространстве и отрицательного давления из-за эффекта Бернулли заставляет складки смыкаться-размыкаться, т.е. обеспечивает нормальный режим их колебаний (рисунок 6). При этом происходит модуляция потока воздуха, который порциями (как в духовых инструментах) вталкивается в резонансные полости. Последовательность воздушных толчков, возникающих в результате колебаний голосовых связок, называется глоттальной волной, обычно она представляется в виде зависимости объемной скорости воздуха от времени (рисунок 7). Как видно из графиков, такой сигнал представляет собой последовательность импульсов, форма которых зависит от соотношения времени открытия складок (скорость потока постепенно нарастает) и времени их закрытия (скорость быстро уменьшается). Период такой волны определяется длительностью общего цикла колебаний связок, т.е. основной частотой колебания. Амплитуда определяется максимальной скоростью потока воздуха, которая, в свою очередь, зависит от величины подскладочного избыточного давления.

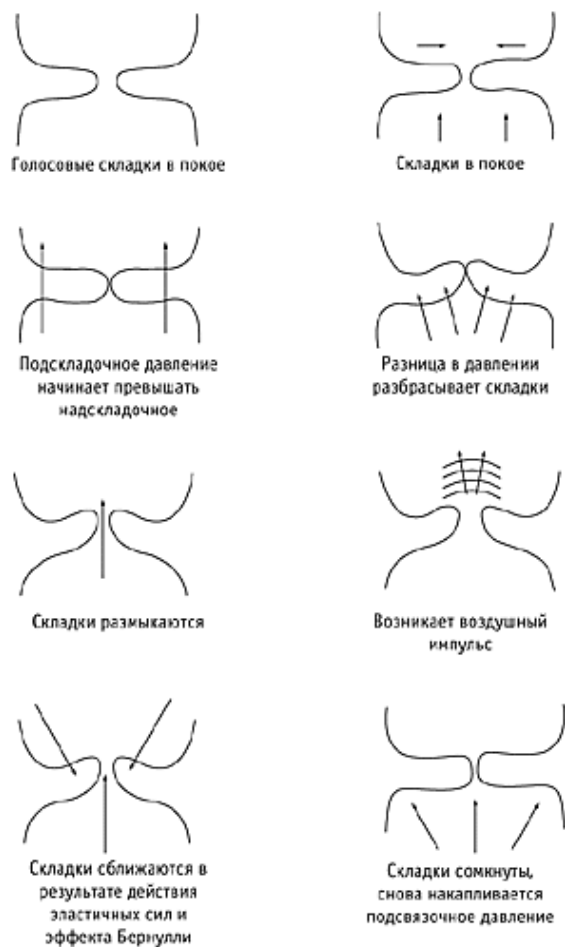


Рис. 6
Процесс колебаний голосовых складок

Частота колебаний складок определяет *высоту голоса* (у мужских голосов при речи она равна в среднем 110 Гц, у женских – 220 Гц), *амплитуда* определяет его *громкость*.

Если записать микрофоном такой звук у самых голосовых складок, то он напоминает гудение или жужжание. Это как бы исходный материал – чтобы получить из него звуки речи, его еще надо обработать в

артикуляционном тракте. Спектр такого звука показан на рисунке 7. Поскольку колебания голосовых складок создают периодический сигнал (реальный сигнал не является строго периодическим), то спектр его при нормальной фонации является гармоническим с крутизной убывания 12 дБ/окт. Для увеличения громкости речи необходимо увеличить подскладочное давление (затратить больше энергии), при этом фронты голосовых импульсов становятся более крутыми (складки быстрее открываются). Время, когда щель закрыта, увеличивается от 40...50% при нормальной фонации, до 65...70% – спектр соответственно изменяется, в нем появляется больше гармоник, что соответственно меняет тембр голоса (делает его ярче).

Способы смыкания складок при фонации могут быть разными. Например, если складки смыкаются не полностью, и между ними имеется щель, то форма импульсов становится почти симметричной, скорость не падает до нуля, в голосе слышен шум (придыхательный голос, шепот). Наоборот, если складки слишком сильно смыкаются (голос становится зажатым), это также меняет форму импульсов и, соответственно, спектр и тембр голоса.

Все перечисленные характеристики – основная частота колебаний голосовых связок, форма голосовых импульсов, их амплитуда, спектральный состав и форма огибающей спектра – играют существенную роль при слуховом восприятии речи. Особую роль играет частота основного тона: в речевом потоке она определяет высоту голоса, и ее изменение используется также для изменения интонации, логических ударений, а иногда и смысла слов (например, в тональных языках, таких, как китайский). В вокальной речи (пении) частота основного тона может изменяться в широких пределах, обычно одна-две октавы (хотя были уникальные певцы с возможностью изменения высоты основного тона до четырех октав – Има Сумак, Мадо Робен и др.).

Частота основного тона, т.е. число колебаний голосовых связок в секунду, зависит от их длины, массы и натяжения. Приблизительно эту связь можно представить, как для струны (хотя они больше похожи на резиновые шнуры) в виде: $f=0,5\sqrt{T/LM}$, где T – натяжение (упругость), L – длина, M – поверхностная масса. Таким образом, чем длиннее и тяжелее складки (эти свойства врожденные), тем более низкий тон имеет голос, чем короче и тоньше, – тем голос выше. Масса зависит от длины, толщины и плотности складок. В процессе речи и пения толщина и плотность складок может значительно меняться за счет натяжения.

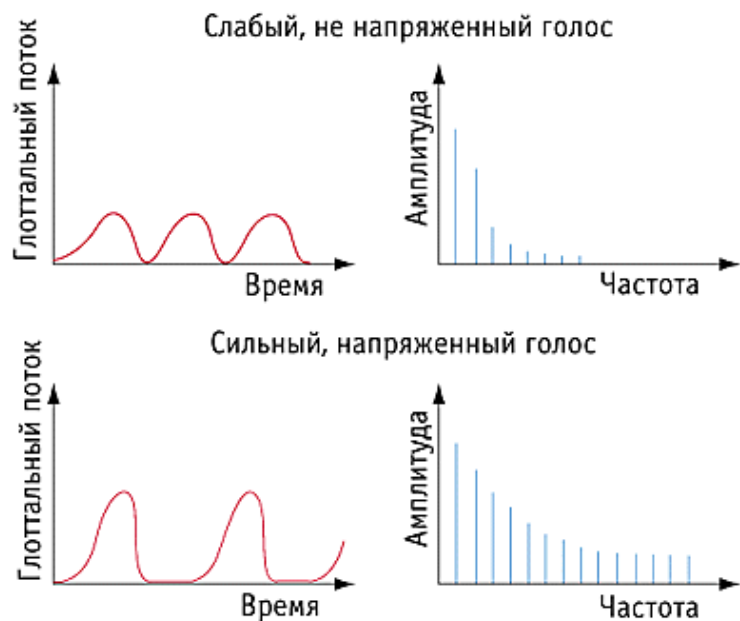


Рис. 7
Форма голосовых импульсов и их спектр

Натяжение обеспечивает повышение высоты голоса, и может осуществляться за счет напряжения внутренних вокальных мышц (в основном при речи) и поворота щитовидного и перстневидного хрящей относительно друг друга (в основном при пении). Поскольку при увеличении громкости голоса растет подскладочное давление, а оно также оказывает некоторое влияние на натяжение складок (мышцы рефлекторно напрягаются), то обычно, при повышении громкости речи растет и высота тона (например, при крике). Только специально обученные певцы могут удерживать высоту тона при увеличении громкости в определенных пределах.

Таким образом, при образовании звуков речи с помощью процесса фонации (т.е. колебания голосовых связок) формируется звуковой сигнал, который затем трансформируется в вокальном тракте, где он превращается из "сырого" материала в последовательность речевых акустических сигналов (другие способы создания источников звука будут рассмотрены позднее).

Таким образом, вокальный тракт выполняет функцию резонатора, т.е. усиливает и фильтрует входной сигнал (аналогично трубам духовых инструментов). Форма труб вокального тракта показана на рисунке 8. Как видно из рисунка, тракт состоит из трех основных резонансных полостей: глотка, ротовая полость, носовая полость. Схематически его вид показан на рисунке 8. Отличия такой системы резонаторов от любых труб в

музыкальных инструментах заключаются в следующем:

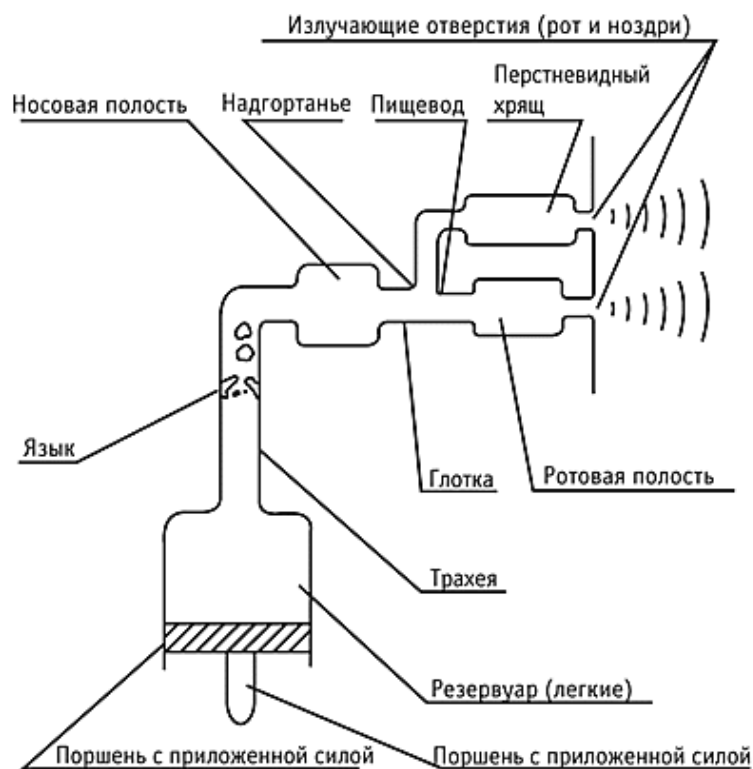


Рис. 8

Схематическая модель вокального тракта

- сложная геометрическая форма: вокальный тракт можно рассматривать как трубу переменного сечения с подключением параллельной трубы (носовой полости, которая может подключаться при опускании заднего мягкого язычка);

- возможность быстрой перестройки формы труб, площади их поперечного сечения, плотности и жесткости стенок, за счет изменения положения языка, мягкого язычка, губ, зубов, расширения глотки, опускания гортани и др. Возможности перестройки параметров вокального тракта огромны, присущи только человеку, что и позволяет ему произносить все многообразие звуков речи. Этот процесс перестройки называется **артикуляцией**. Каждому звуку речи соответствует либо определенное статическое положение, либо определенная динамика изменения положения языка, челюстей, губ, нёбной занавески, т.е. определенная артикуляция.

Общая длина речевого тракта у взрослого человека (от голосовых складок до губ) около 17 см, длина носовой полости (от нёбной занавески до ноздрей) 12,5 см, площадь переменного сечения тракта в среднем составляет примерно 5...6 см².

Простейшей моделью вокального тракта можно считать цилиндрическую трубу длиной 17 см, закрытую на одном конце (аналогично трубе кларнета). Собственные моды (формы) колебаний такой трубы показаны на рисунке 9, частоты определяются из соотношений: $L = \lambda/4$; $L = 3\lambda/4$; $L = 5\lambda/4$ и т.д., таким образом частоты равны $f_n = (2n-1)c/4L$, где n -целое число; L -длина трубы; c -скорость звука.

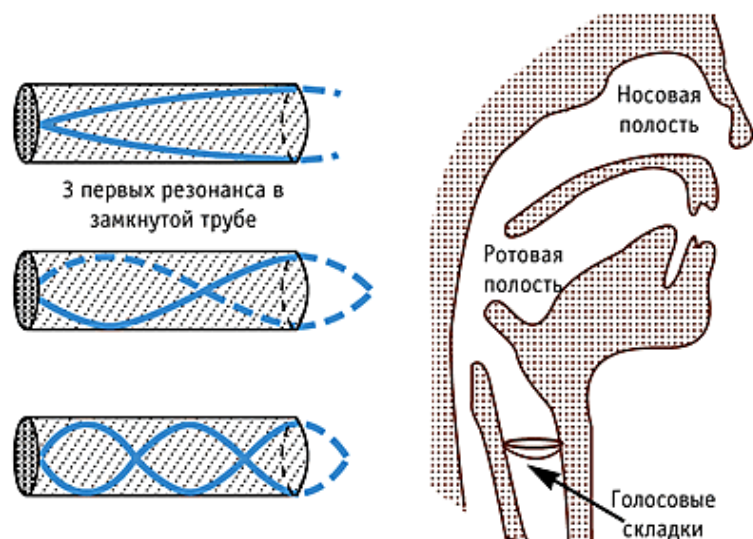


Рис. 9
 Формы колебаний для цилиндрической трубы
 и голосового тракта

В спектре такой трубы присутствуют только нечетные гармоники 1:3:5... Для длины $L = 17$ см, собственные частоты оказываются равными 500, 1500, 2500 Гц. Если у трубы менять в разных точках площадь поперечного сечения, то положение ее собственных частот будет смещаться. Совершенно аналогичные процессы происходят в вокальном тракте: в нем также имеется свой набор собственных частот с соответствующими модами колебаний, т. е. определенным распределением узлов и пучностей вдоль его длины. Меняя площадь поперечного сечения в вокальном тракте, можно также все время менять положение собственных частот.

Если на вход такой трубы (системы труб) подать сигнал, сформированный при колебаниях голосовых связок (рисунок 7), то на выходе можно записать сигнал, который будет иметь форму, показанную на рисунке 10, т.е. гармоники, совпадающие с собственными частотами тракта, будут усилены за счет резонансов.

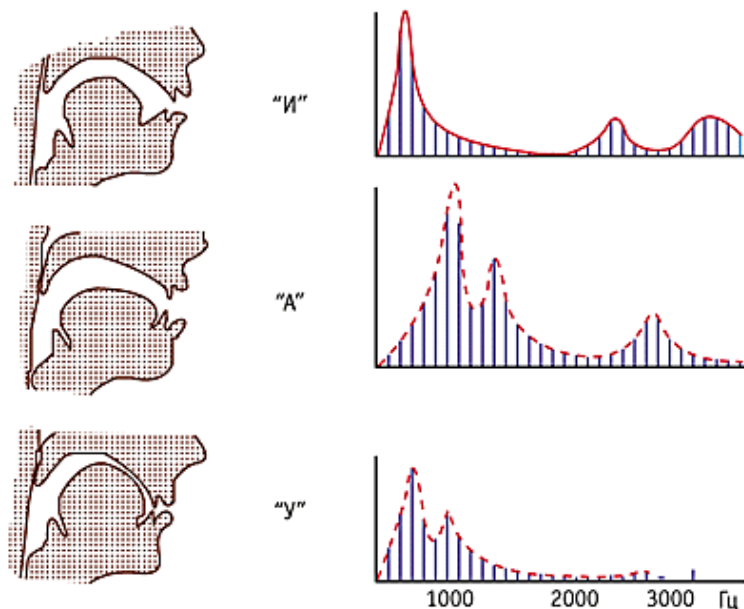


Рис. 10
 Положение тракта для разных звуков речи
 и вид звукового сигнала с формантами

Области спектральных максимумов, соответствующие резонансным частотам вокального тракта, называются **формантами** (иногда их просто называют резонансами вокального тракта). Каждому звуку речи (простейший звук речи называется фонемой) соответствует своя форма вокального тракта, которая варьируется за счет изменения положения языка, губ, зубов и т.д., и свое положение формант (F-картина). Примеры показаны на рисунке 10.

Существуют некоторые общие закономерности в управлении расположением собственных частот резонаторов: если поперечное сечение трубы уменьшается в области, где форма колебаний (мода), соответствующая данной резонансной частоте (форманте), имеет максимум давления, то частота увеличивается; если в точке, где минимум давления, то частота уменьшается. Изучение движения артикуляционных органов во время речи с помощью рентгенографических съемок показали, что аналогичные закономерности имеют место и в вокальном тракте: при подъеме языка вперед и вверх сужается передняя часть ротовой полости, при этом понижается первая форманта F_1 и повышается вторая F_2 . При сдвиге языка назад сужается поперечное сечение тракта в области глотки, при этом повышается F_1 и понижается F_2 и т.д. При сдвиге формант по определенным закономерностям происходят изменения в соотношении их амплитуд, что приводит к изменению формы огибающей. Все эти признаки (расположение формант и соотношение их амплитуд) и являются отличительными акустическими признаками гласных звуков речи.

Правда, при беглой речи происходит настолько быстрая перестройка позиции артикуляционных органов (языка, губ и др.), что часто имеет место наложение позиции, соответствующей одному звуку, на позицию другого (обычно гласного на соседний согласный), такое явление называется *коартикуляцией*, и оно очень осложняет восприятие и распознавание речи.

Таким образом, вокальный тракт действует на звуковой сигнал источника как параметрический эквалайзер, при этом существенное значение имеют частоты резонансов, соотношения их амплитуд и ширина резонансных пиков (добротность). Примерные области расположения первых трех формант для гласных русского языка даны в таблице.

Частотный диапазон формант (Гц)		Ширина формант (Гц)
Тип голоса	Мужской	Женский

F1	200-800	250-1000	40-70
F2	600-2800	700-3300	50-90
F3	1300-3400	1500-4000	60-180

Распознавание каждой фонемы происходит в основном по положению первых двух формант F1 и F2, более высокие форманты определяют тембральные различия (для пения чрезвычайно существенное значение имеет третья формантная область "певческая форманта"). Расположение формант для гласных английского языка показано на рисунке 11.

Если подходить к процессу образования звуков речи с помощью фонации в терминах передаточных функций, то этот процесс может быть описан следующим образом: $P(\omega) = S(\omega)T(\omega)R(\omega)$, где $S(\omega)$ – передаточная функция входного сигнала, $T(\omega)$ – передаточная функция тракта, $R(\omega)$ – активная составляющая сопротивления излучения, (рисунок 12). Именно эта последовательность операций и реализуется в различных синтезаторах звука. Под передаточной функцией тракта понимается отношение комплексных амплитуд объемной скорости на губах U_0 к объемной скорости у голосовой щели U_r : $T(\omega) = U_0/U_r$. Для цилиндрической трубы с одним закрытым концом она вычисляется по формуле: $T(\omega) = 1/\cos(2\pi fLc)$. На резонансных частотах, определяемых по формуле $f_n = (2n-1)c/4L$, знаменатель обращается в нуль, и функция имеет максимумы (из-за наличия затухания она имеет конечные значения).

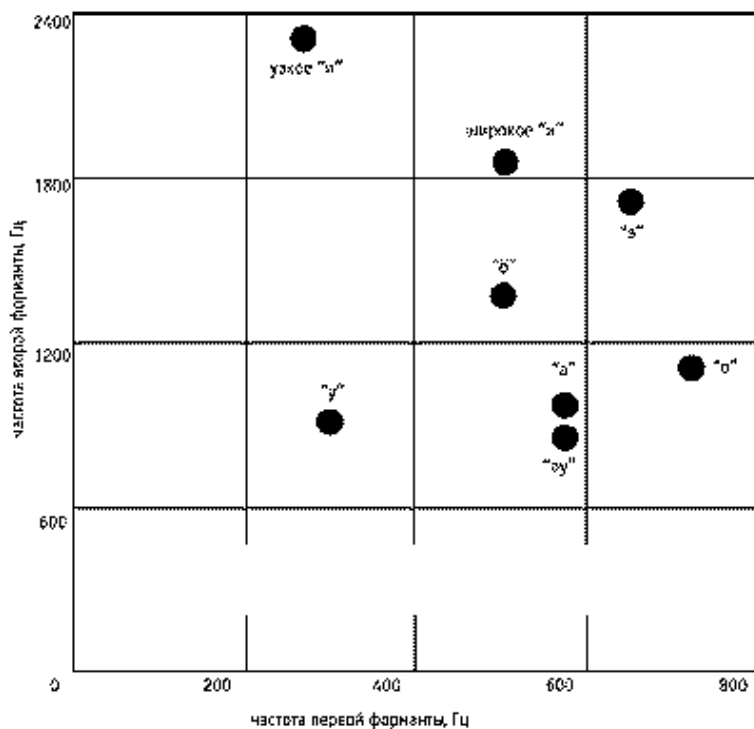


Рис. 11
Форманты для гласных

В реальном голосовом тракте передаточная функция имеет более сложный характер (она может быть вычислена и измерена современными цифровыми методами), но на резонансных частотах тракта, т.е. на формантах, она также имеет максимумы, которые называются полюсами. Таким образом, форманты еще можно определить как полюса передаточной функции.

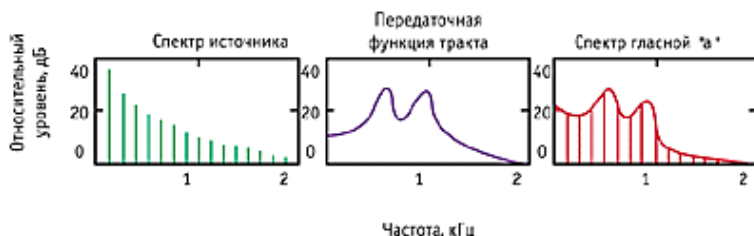


Рис. 12
Передаточная функция голосового тракта

Описанные выше процессы голосообразования относятся в основном к гласным звукам, процессы образования согласных звуков существенно сложнее, и будут рассмотрены в следующей части статьи.

Часть 17. Слух и речь. Часть 2

Образование согласных. Классификация звуков речи

Как уже было сказано в первой части статьи (1/2002), при образовании звуков речи используются три основных механизма генерации звука (и их различные сочетания): процесс фонации, турбулентный шум, импульсные источники.

Процесс фонации, т.е. модуляция воздушного потока за счет колебаний голосовых связок, что используется при образовании гласных звуков и звонких согласных, формирует последовательность импульсов которые затем подвергаются фильтрации в голосовом тракте. В результате этого формируется речевой акустический сигнал.

Турбулентный шум проявляется при прохождении струи воздуха через узкое отверстие с достаточно большой скоростью. В струе может происходить вихреобразование, сопровождающееся турбулентным шумом (рисунок 1). Такой способ звукообразования используется в лабиальных музыкальных инструментах (например, флейте), этот же способ звукообразования используется и в голосовом тракте: в определенном месте тракта создается сужение, и при продувании воздуха через него образуется шум, который затем определенным образом трансформируется в резонансных полостях тракта.

Спектр турбулентного шума имеет плоский участок в диапазоне 500...3000 Гц, выше и ниже которого он спадает со скоростью -6 дБ/окт. Пример спектрального распределения шума для звуков "ф", "с", "ш", "х" показан на рисунке 2.

Импульсный источник звука образуется при скачкообразном изменении давления воздуха. Для образования звукового импульса необходимо создать в речевом тракте значительное избыточное давление, полностью перекрыв выход воздуха на некоторый интервал времени (примерно 30 мс). Амплитуда такого импульса зависит от величины давления позади смычки, поэтому в речи более громкими будут те согласные, место скачка давления для которых находится ближе к гортани – там давление накапливается быстрее, и его уровень выше. Импульсный звук такого типа имеет сплошной спектр.

Каждый из трех перечисленных выше способов (фонация, турбулентный шум, импульс) может выступать как самостоятельный источник при образовании звуков речи, а может использоваться в различных сочетаниях.

Образование согласных звуков

В отличие от гласных звуков согласные отличаются гораздо большим разнообразием способов звукообразования, что усложняет их анализ и распознавание, и соответственно затрудняет работу звукорежиссеров с речевыми и вокальными сигналами. Однако именно согласные требуют особого внимания при звуковой обработке, поскольку в речи они несут основную смысловую нагрузку.

Во-первых, при образовании согласных могут использоваться все источники звука и их сочетания: фонация, турбулентный шум, импульс.

Во-вторых, месторасположение источника звука также может сильно варьироваться: если при образовании гласных резонаторы всегда находятся впереди источника звука, поскольку положение голосовых связок закреплено в гортани, то при образовании согласных источник звука может находиться в любом месте тракта (например, у зубов для звука "с"; в задней нёбной части для звуков "г", "к" и др.).

В-третьих, при образовании согласных чаще используется подключение дополнительной носовой полости (в русском языке вообще нет носовых гласных, только согласные "м", "н").

Кроме того, они отличаются значительно более короткими стационарными периодами (служат как бы переходом от одной гласной к другой), и значительно большим разнообразием спектров. Средняя длительность гласных звуков 0,15 с, средняя длительность согласных 0,08 с.

При создании согласных звуков процесс образования формантных областей значительно усложняется.

Поскольку источник возбуждения (вибратор) может располагаться в любом месте голосового тракта, то в этом случае резонансные полости располагаются как перед источником, так и позади него.

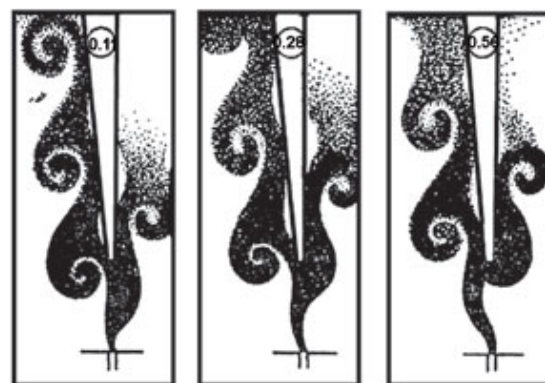


Рис. 1 Процесс вихреобразования в турбулентной воздушной струе

Резонансы полости перед источником создают пики в выходном сигнале, такие резонансы называются "полюсами" (формантами); резонансы задних полостей называются "нулями" передаточной функции, и проявляются в виде провалов. Модель тракта для согласных "к", "п", "т" и соответствующая передаточная функция тракта показана на рисунке 3. На графике отчетливо видны пики (полюса) и провалы (нули).

Когда нули и полюса находятся близко к друг к другу, происходит их нейтрализация, и на выходной характеристике не видно ни нулей, ни полюсов. В таком случае они называются "связанными".

Для описания процессов образования согласных звуков вводится понятие *локусной формантной картины*. При образовании гласных звуков передаточная функция тракта, как уже было отмечено выше, полностью зависит от структуры ее формант (F-картины), которые в процессе беглой речи непрерывно изменяются, поскольку непрерывно перестраиваются артикуляционные органы. Эта плавность изменения конфигурации речевого тракта и его резонансных частот имеет место и при произнесении согласных звуков, только эти резонансы не всегда видны в передаточной функции.

Под локусной F-картиной понимается совокупность резонансов (формант) ротовой полости тракта, которая соответствует положению артикуляционных органов при произнесении данного согласного звука. Таким образом, локусы – это те форманты, которые должны быть при данной конфигурации тракта, независимо от того, слышны они или нет. Их положение можно восстановить из спектрограмм сигнала, и оно имеет существенное значение для процессов восприятия согласных.

Артикуляционные возможности речевого тракта при образовании звуков чрезвычайно разнообразны, и могут быть использованы для создания огромного многообразия звуков. Однако для речи используется ограниченный набор звуков (количество фонем в разных языках мира в основном не превышает 50...70). Такой разрыв между возможностями голосового аппарата и его применением объясняется с помощью квантальной теории, в соответствии с которой из всех звуков в речи используются только те, которые создают достаточно четкие слуховые контрасты и легко различимы слуховой системой (т.е. речь была приспособлена к слуху). Например, гласные "и", "у", "а" резко контрастируют на слух, поэтому они используются почти во всех языках мира. Поэтому для разных звуков для речи были отобраны те виды артикуляции, которые создают существенные акустические и слуховые различия.

Рассмотренные выше механизмы звукообразования с учетом квантальных артикуляционно-акустических и слуховых отношений и лежат в основе классификации звуков речи, краткое ознакомление с которой необходимо для анализа акустических характеристик речи и их связи с фонетическими признаками в процессе слухового восприятия и распознавания.

Классификация звуков речи

В основе классификации всех звуков речи, участвующих в различении слов, лежит классификация Международной фонетической ассоциации (МФА), основанная на артикуляционных признаках. Однако, поскольку все звуки, с одной стороны, представляют собой семиотические знаки, создаваемые органами речи, а, с другой, они представляют собой акустические сигналы, восприятие которых создает сложные слуховые образы, существуют классификации, основанные на акустических признаках. Признаки, используемые в классификации МФА, делятся на группы, в зависимости от того, какой из нижеперечисленных процессов они описывают:

- способы формирования воздушного потока – генерации (инициации);
- способы участия голосовых связок в образовании звука – фонации;
- способы формирования структуры вокального тракта – артикуляции.

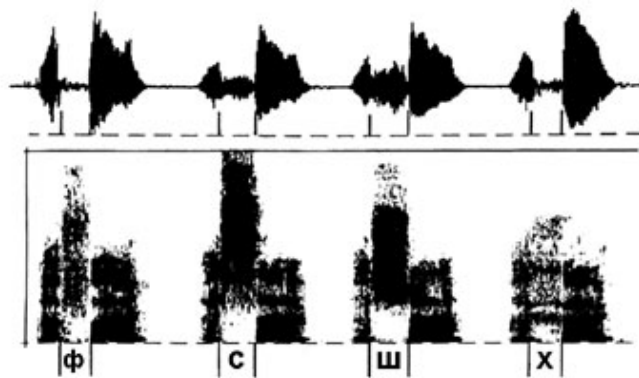


Рис. 2 Осциллограммы и спектрограммы шума при образовании глухих согласных

Все звуки речи можно разделить на две большие группы: гласные и согласные, существенно отличающихся друг от друга по всем вышеперечисленным признакам.

Поскольку согласные звуки несут основную смысловую нагрузку в тексте речи, а гласные – основную эмоциональную нагрузку, то разное количественное сочетание этих звуков в разных языках определяет различие в избыточности речи, и разные требования к их обработке в каналах звукопередачи и звукозаписи.

Классификация гласных

При образовании гласных звуков всегда используется один способ формирования воздушного потока (генерации или инициации): модуляция потока воздуха за счет колебаний голосовых связок (фонация), поэтому этот признак не может использоваться для их классификации.

В основе классификации гласных лежат другие признаки.

Дополнительная тембровая окраска гласных, которая используется в разговорной и вокальной речи требует сложной вокальной артикуляции с дополнительными движениями, модифицирующими свойства вокального тракта. Наконец, при создании гласных могут использоваться такие признаки как:

- долгота (долгий/короткий) – в русском языке они не несут смыслового различия, а в других языках, например, английском, это существенный различительный признак.
- напряженность (напряженный/ненапряженный) – этот различительный признак также используется в ряде языков. Ненапряженные гласные отличаются меньшей длительностью и интенсивностью, некоторым сдвигом в артикуляции.

Кроме чистых гласных, во многих языках используются сложные гласные, где происходит плавный переход от одного типа артикуляции к другому. Если в этом участвуют два гласных, то такой звук называется дифтонгом. Например, в русском языке: "я" [йа], "ю" [йу]. Существуют сочетания, где используются три звука (трифтонги).

Классификация согласных

Артикуляция согласных звуков связана с созданием препятствия на пути воздушного потока в различных частях голосового тракта. Кроме того, при образовании согласных используются все три типа генерации (инициации) звука: фонация, турбулентный шум и звуковой импульс (взрыв) и их всевозможные сочетания. Поэтому классификация согласных осуществляется по всем трем вышеперечисленным критериям.

По способам генерации

Основной способ создания воздушного потока у большинства согласных – легочный выдыхательный механизм (как и у гласных).

По способам артикуляции

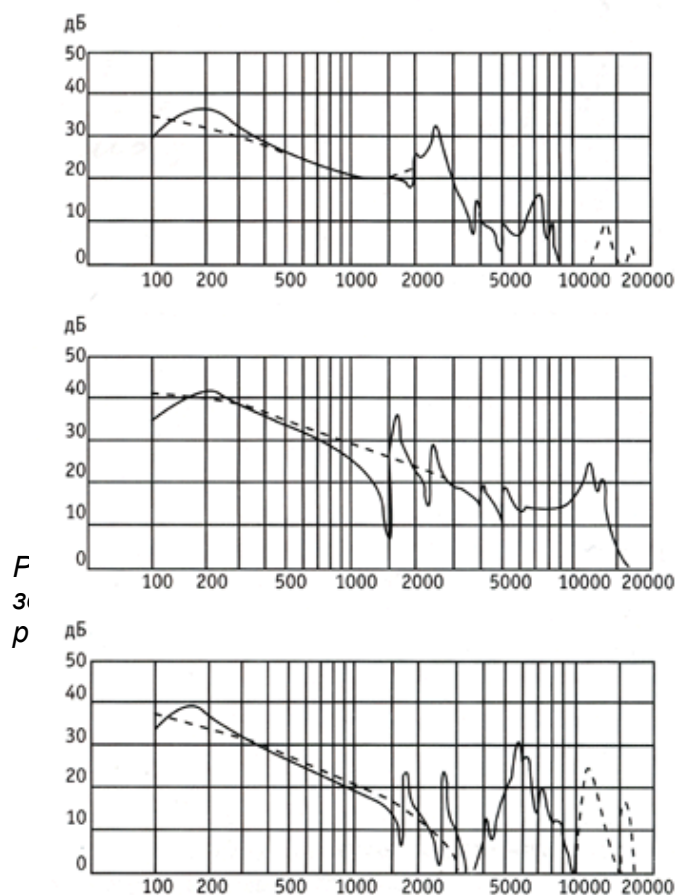
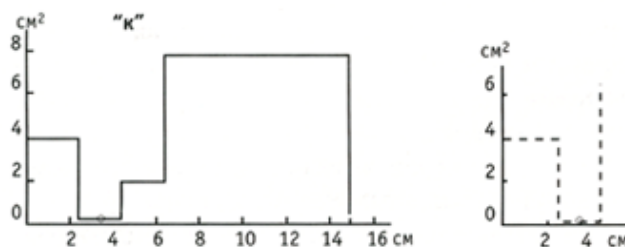
При классификации по этим способам используются следующие основные признаки: способ образования преграды, и место ее образования.

Образование согласных может сопровождаться дополнительными сложными артикуляционными движениями. Например, можно выделить согласные, которые образуются с образованием *двойной преграды* (их называют иногда двухфокусными) – "ш", "ж"; если эта вторая преграда образуется за счет сближения губ, то такие согласные называются лабиальными, например, английское "w".

Кроме того, имеются согласные с *дополнительной артикуляцией*, например, широко распространенные в русском языке мягкие согласные "л'", "м'", которые образуются путем наложения дополнительной язычной артикуляции [и]-образного типа; этот процесс называется палатализацией.

Наконец, согласные могут различаться по длительности (*долгий/короткий*) и степени напряженности артикуляции (*сильный/слабый*). В русском языке они не несут смысловых различительных функций.

Кроме того, все согласные, при образовании которых не образуется сильного турбулентного шума, потому что при их образовании имеется дополнительный проход для воздуха, объединяются под названием "*сопоровые*". К их числу относятся носовые согласные ("м", "н"), аппроксиматы ("л", "р") и полугласные "й".



Все вышесказанное относится к отдельным звукам, но в беглой речи имеет место процесс *коартикуляции*. В процессе артикуляции любого звука имеется три фазы:

- подготовительная (экскурс), когда органы речи начинают устанавливаться в исходную позицию;
- стационарную, когда все органы (язык, губы и др.) находятся в точной позиции, соответствующей данному звуку;
- рекурсия, когда органы речи начинают перестройку для следующего звука.

Если бы для каждого звука все эти позиции точно выдерживались, как при четком произношении отдельного звука, то речь происходила бы в слишком медленном темпе. При быстрой речи (14...18 звуков в секунду), происходит переслаивание этих фаз между соседними звуками. Этот процесс влияния артикуляции соседних звуков друг на друга называется коартикуляцией. При этом на артикуляционное положение органов речи для данного звука накладываются положения (движения) органов речи, соответствующие последующему звуку.

Эти процессы коартикуляции в беглой речи существенно влияют на акустические характеристики речи и процессы ее слухового распознавания, о чем мы поговорим в следующих статьях.

Часть 17. Слух и речь Часть 3. Акустические характеристики речи

Речевой сигнал имеет двойственную природу – с одной стороны, это обычный акустический сигнал, который представляет собой процесс распространения энергии акустических колебаний в упругой среде. Как любой акустический сигнал, он может быть представлен в виде звуковых волн, представляющих собой распространение процессов сжатия и разряжения частиц среды, формы фронтов которых зависят от свойств источника и условий распространения. Поэтому, как и другие акустические сигналы, речь характеризуется определенным набором объективных характеристик: зависимостью звукового давления от времени (временной структурой звуковой волны), длительностью звучания, спектральным составом, местом расположения источника в пространстве и пр.

С другой стороны, речь как физическое явление вызывает определенные субъективные слуховые ощущения (громкости, высоты, тембра, локализации, маскировки и др.), именно проблемам их взаимодействия и были посвящены предыдущие статьи по психоакустике.

Речевой сигнал подвергается такой же процедуре обработки в слуховой системе, как и любой другой акустический сигнал, т. е. на основе его анализа формируются те же слуховые ощущения – например, восприятие речи на абсолютно незнакомом языке ничем не отличается от восприятия окружающей акустической информации – шума, свиста, щелчков и др. Однако если человек воспринимает речь на языке, которому он был предварительно обучен, то наряду с обработкой чисто акустической информации (громкости, высоты, тембра и пр.) происходит фонетическая, а вслед за ней и семантическая расшифровка информации, для чего подключаются специальные отделы головного мозга.

На протяжении уже многих десятилетий, и особенно интенсивно в последние годы, в связи с развитием технологии и систем автоматического распознавания и синтеза речи, изучаются акустические характеристики речевых сигналов, и предпринимаются попытки установления связи между акустическими параметрами и фонетическими признаками речевых сигналов, т.е. попытки понять, как мозг, получив информацию о характере изменения звукового давления во времени, извлекает информацию о смысловом содержании речи. В этом направлении получено уже очень много результатов: количество книг и статей по этим проблемам исчисляется тысячами, в качестве примера могу привести одну из последних книг знаменитого ученого М. Шредера "Computer Speech: Recognition, Compression, Synthesis" (Берлин, 1999 г.).

Однако изучение чисто акустических характеристик речевых сигналов представляет значительную самостоятельную ценность для систем звукозаписи, радиовещания, компьютерной обработки речи и др., т.е. для всех процессов записи, обработки, передачи и воспроизведения речевых сигналов, которые принципиально важны для работы звукорежиссера. Поэтому начнем с анализа акустических характеристик речевых сигналов, а затем попробуем остановиться на их связи с фонетическими признаками, и на существующих в настоящее время теориях слухового восприятия и обработки речи.

Анализ акустических характеристик речевого сигнала начинается с записи изменения звукового давления во времени с помощью микрофона – эта зависимость мгновенного значения звукового давления от времени представляется в виде осциллограммы. Обычно в технических приложениях, в частности при компьютерной обработке, происходит запись усредненного за некоторый отрезок времени уровня звукового давления от времени, эта зависимость называется *уровнеграммой*. Пример уровнеграммы слова "welcome" показан на рисунке 1.



Рис. 1. Уровнеграмма речевого сигнала

Вид уровнеграммы существенно зависит от времени и способа усреднения – во всех звуковых программах об этом спрашивается пользователь, (правда, как показывает практика, он не всегда об этом догадывается).

Способ усреднения может быть равномерный или экспоненциальный (например, uniform или exponent в программе Sound Forge). Обычно выбирается время усреднения для пиковой уровнеграммы 1...2 мс, для объективной 15...20 мс, и для субъективной 150...200 мс. В первом случае получается точная запись пиковых значений сигнала; во втором отсутствуют излишние мелкие детали (это время обычно используется при компьютерной обработке речи); наконец, в последнем выбрано время, в течение которого слуховая система опознает тембр.

Если средние значения сигналов сохраняются равными на определенных отрезках времени, то такие сигналы называются стационарными. Звуковые сигналы (речевые и музыкальные) являются сигналами квазислучайными и нестационарными, хотя для речи можно указать приближенно такие отрезки времени (порядка 2...3 мин), при которых речевые сигналы можно считать квазистационарными.

Полученные уровнеграммы позволяют провести статистический, корреляционный и спектральный анализы речевого сигнала, что можно делать с помощью обычных аудиопрограмм, а также с помощью специальных программ, предназначенных именно для речевых сигналов с учетом их специфики: Ultra- sound (Австралия), CSRE (Англия), Viper (Германия), Praat (Голландия), Phonograph (Россия) и др.

Поскольку речевой сигнал, как и музыкальный, представляет собой сигнал квазислучайный, т.е. предсказать его будущие значения можно только с определенной вероятностью, то для анализа его характеристик могут быть применены все известные методы статистического анализа. При этом исследуется распределение во времени следующих величин:

- мгновенных значений и уровней речевого сигнала;
- длительностей непрерывного существования разных уровней;
- длительностей пауз;
- распределение максимальных уровней по частоте;
- распределение текущей и средней мощности;
- спектральной плотности мощности.

Кроме того, могут быть определены такие важные для практики звукозаписи параметры, как динамический диапазон и пик-фактор, вычислено распределение основной фонационной частоты, спектральное распределение формант и др.

Знание статистических характеристик речевых сигналов необходимо для оптимальной организации систем звукового вещания, систем звукозаписи, современных систем сжатия речевого сигнала и др. Исследование этих характеристик для русской речи было выполнено в работах Фурдуева, Римского-Корсакова, Сапожкова, Белкина, Шитова и др.

Непосредственно из анализа уровнеграмм речевого сигнала прежде всего может быть получена информация о распределении мгновенных значений и уровней звукового сигнала во времени, и длительности их превышения установленного значения. Это позволяет определить динамический диапазон и пик-фактор речевого сигнала, а также установить распределение длительности пауз, отрезков непрерывных речевых звучаний, распределения текущей и средней мощности сигнала во времени и др.

Если очень коротко остановиться на этих данных, то можно отметить, что распределение плотности вероятности мгновенных значений речевого сигнала, показанное на рисунке 2, носит экспоненциальный характер, и существенно отличается от нормального распределения, которому подчиняется например, джазовая или хоровая музыка. Статистический анализ длительности непрерывного существования разных уровней в речевом сигнале показывает, что наиболее вероятными являются выбросы (пики) длительностью 12...17 мс, из чего следует, что максимальные уровни сигнала достигаются в кратковременные промежутки времени.

Распределение длительности пауз в речевых сигналах также носит случайный характер, их средняя длительность для речи составляет 0,4 с, а суммарная длительность пауз достигает 5% от времени передачи. Наиболее важная информация, которую позволяет получить анализ уровнеграмм – это определение динамического диапазона речевого сигнала и его пик-фактора. Динамическим диапазоном звукового сигнала называется разница между его квазимаксимальным и квазимиимальным уровнем $D = L_{max} - L_{min}$. Под квазимаксимальным L_{max} понимается такой уровень сигнала, длительность пиков выше которого составляет 1% (для речи) и 2% (для музыки) от общей длительности отрезка сигнала. Аналогично определяется квазимиимальный уровень L_{min} (относительная длительность составляет 99% и 98%). Значения пик-фактора определяются как разница между квазимаксимальным и средним уровнем сигнала $D = L_{max} - L_{cp}$.

Значения динамических диапазонов речевых сигналов находится в пределах 35...45 дБ, значения пик-фактора 10...12 дБ.

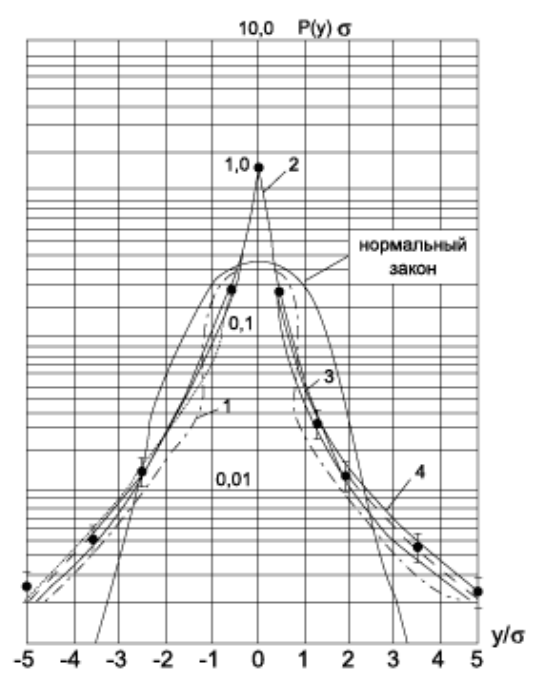


Рис. 2. Распределение плотности вероятности мгновенных значений речевого сигнала. 1 – дикторский текст; 2, 3, 4 – художественное чтение

Условия	Расстояние (см)	Среднее звуковое давление, Па (дБ)	Пиковое значение мощности (мВт)	Пик-фактор (дБ)	Область максимальных уровней (Гц)
Речь телефонная	2,5	2 (100)	0,24	12	250-500
средний уровень		4 (106)	4	18	500-1000
громкий		1 (94)	0,025	8	250-500
тихий					
Разговор	100	0,05 (68)	0,5	10	250-500
Оратор	100	0.1 (74)	2.0	12	250-500

Некоторые данные для речевого сигнала по развиваемым уровням звукового давления и мощности приведены в таблице.

Если пересчитать уровни звукового давления для телефонной речи на расстояние 100 см, то получатся следующие значения: 68, 74, 62 дБ.

Следует отметить ,что для вокальной речи (пения) эти уровни существенно выше, и могут достигать значений 115 дБ на 1 м. В старом итальянском руководстве по подготовке певцов было написано, что если певец может развивать уровень от 110 дБ и выше, то он может петь в "Ла Скала", если ниже 100 дБ, то в камерном ансамбле, если ниже 90 дБ, то не надо петь вообще... Интересно, сколько народу осталось бы петь на эстраде сегодня при таком критерии?

Корреляционный анализ речевого сигнала позволяет рассчитать функцию текущей автокорреляции и установить предел однородности, которые определяются временем, в течение которого функция автокорреляции достигает некоторого предельного значения, независимого от времени запаздывания. Для речи этот предел составляет 3...5 с.

Спектральный анализ речевого сигнала, как всякого непрерывно изменяющегося во времени акустического сигнала, может быть выполнен на основе записанной уровнеграммы с помощью преобразования Фурье. В любом музыкальном редакторе предусмотрена операция быстрого преобразования Фурье (БПФ, FFT), позволяющая из выделенного отрезка уровнеграммы рассчитать его спектр.

Анализ спектров речевых сигналов позволяет установить форму огибающей и выделить области формантных частот. Поскольку место и ширина формантных областей принципиально важны для распознавания речи, то для точного определения формантных полос в речевом сигнале созданы специальные программы на основе метода линейного предсказания или кепстрального анализа, позволяющие производить их автоматическое

распознавание.

Кроме того, поскольку интонация речевого высказывания определяется изменением частоты фонации, то выделение основной частоты фонации из записанных уровнеграмм и характер ее зависимости от времени имеют принципиально важное значение.

Для интегральной оценки свойств речевого сигнала может быть рассчитан спектр мощности и построено распределение спектральной плотности мощности, которая для речевого сигнала показана на рисунке 3, что позволяет установить, что основная энергия речевого сигнала (B) сосредоточена в полосе 250...1000 Гц, спад в сторону высоких частот происходит со скоростью 7 дБ/окт после 500 Гц.

Анализ спектров дает возможность построить очень важную для практики звукозаписи кривую распределения амплитудного состава речи. Пример для диапазона 1000...1400 Гц показан на рисунке 4, (для других диапазонов распределения аналогичные). Кривая распределения показывает, что более 80% в речевом потоке составляют амплитуды с уровнем 45 дБ, и только менее 10% амплитуды с уровнями 70 дБ и выше. Это значит, что при обработке речевых фонограмм стремление "вычистить шумы" может привести к потере значительной части информации, поскольку низкие уровни амплитуд связаны в основном с согласными звуками, а они являются носителями основной смысловой нагрузки в речи.

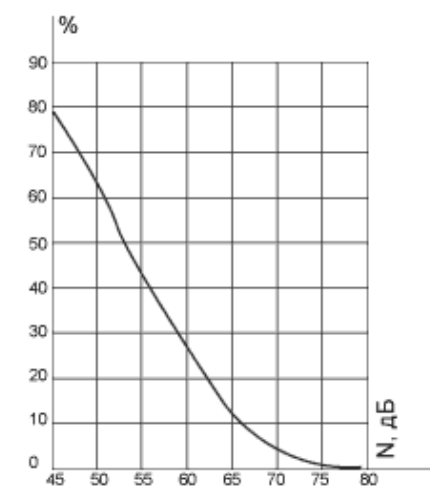


Рис. 4. Амплитудный состав речи в полосе 1000 - 1400 Гц дан на рисунке 5.

Спектрограммы могут быть узкополосные, широкополосные и слуховые. Выбор числа семплов, т.е. выбор длительности отрезка анализируемого сигнала, определяет точность разворачивания по частоте (т.е. расстояние между частотами). Невозможно обеспечить одновременно "хорошее" разворачивание и по частоте и по времени, поскольку они связаны некоторым соотношением $Df \cdot Dt = \text{const}$, (по аналогии с квантовой механикой называемым "принципом неопределенности"). Чем выше точность по частоте, тем хуже разворачивание по времени, и наоборот. Поэтому точность развертки по частоте зависит в обратной пропорции от длительности временного окна при преобразовании Фурье (например, при ширине развертки 100 Гц разворачивание по времени будет $1/100 = 10$ мс).

В практике анализа речевых сигналов применяется два вида спектрограмм: (кумулятивный) спектр речевого широкополосные и узкополосные (рисунки 7а, 7б). В узкополосных спектрограммах используется частота развертки 45 Гц, это ниже, чем самые низкие фонационные частоты в голосе, что позволяет при такой точной развертке отчетливо увидеть вдоль вертикальной оси гармоники голосового источника.

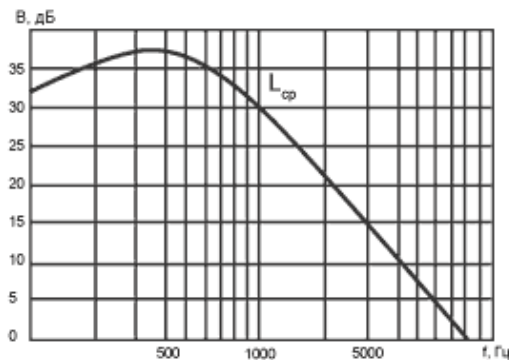


Рис. 3. Спектральное распределение средней мощности речевого сигнала

Кроме одномерных спектров (амплитуда-частота), современные алгоритмы позволяют построить для любого речевого сигнала его трехмерные (кумулятивные) спектры (например, 3D-Frequency Analysis в редакторах Wave-Lab и др.), где по одной оси отложено время, по другой частота, по третьей – амплитуда. (Рисунок 5). Такие спектры позволяют получить значительно больше информации не только о спектральном составе сигнала, но и характере изменения его во времени. Трехмерные спектры широко используются в практике изучения различных акустических сигналов, однако для анализа речевых сигналов наибольшее распространение имеют трехмерные спектры особой формы-спектрограммы.

В 1940 году в лаборатории Bell Lab (США) был построен прибор, получивший название "спектрограф видимой речи", который позволял представить спектр речи в трехмерной форме, только построенной несколько иначе, чем обычный трехмерный спектр. Это своего рода "вид сверху" на трехмерный спектр: по оси абсцисс отложено время, по оси ординат – частота, а амплитуда показана интенсивностью цвета (чем интенсивнее, тем больше амплитуда). На рисунке 6 показан пример спектрограммы того же речевого сигнала, 3D-спектр которого

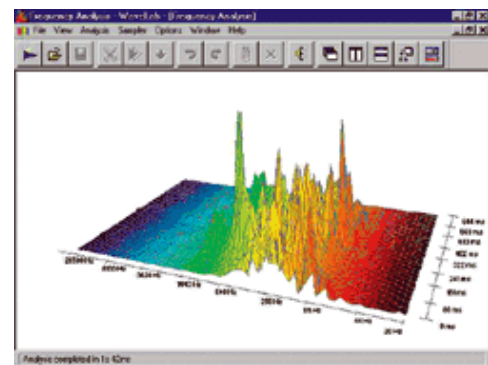


Рис. 5. Трехмерный спектр речевого сигнала

Как было сказано в предыдущих статьях, речевой сигнал – это результат "свертки" (умножения) звукового сигнала, создаваемого голосовым источником, например, за счет модуляции воздуха при колебаниях голосовых связок, и огибающей, за счет резонансных свойств голосового тракта (этим и определяется его формантная структура).

На широкополосных спектрограммах, обычно с частотой развертывания 300 Гц, отчетливо видны вертикальные полосы вдоль оси времени, связанные с появлением отдельных импульсов воздушного давления при колебаниях голосовых связок, и сильно подчеркнуты темные горизонтальные полосы, соответствующие формантам. Поэтому, в зависимости от целей, которые ставятся при анализе речевого сигнала, используются или широкополосные спектрограммы (выделяются отдельные импульсы воздуха, подчеркнуты форманты), или узкополосные, где выделяются обертоны голосового источника. При этом можно проследить изменение основной частоты фонации во времени, что имеет большое значение при оценке мелодического рисунка речи, как было отмечено выше. Кроме того, полученные значения спектров позволяют оценить распределение энергии во времени.

Однако, ни широкополосная, ни узкополосная спектрограммы не учитывают специфику спектрального анализа сигнала, который производится во внутреннем отделе периферической слуховой системы на базилярной мембране. Поэтому в последние годы с учетом новейших результатов в психоакустике была разработана методика построения "слуховых" спектрограмм. При построении этих спектрограмм используются фильтры с различными полосами пропускания, ширина которых соответствует ширине "критических полос" слуха (или ширине слуховых фильтров при спектральном анализе звуков на базилярной мембране).

Ширина критических полос зависит от частоты, эта зависимость примерно соответствует ширине третьоктавных полос. В такой спектрограмме на низких частотах (первые 4...5 критических полос) происходит узкополосная обработка сигнала по частоте. На высоких частотах критические полосы становятся значительно шире, это соответствует широкополосной спектрограмме, т.е. идет очень точное развертывание по времени.

Таким образом, слуховая спектрограмма значительно точнее отражает восприятие и обработку речевого сигнала в слуховой системе: на низких частотах основное внимание концентрируется на отдельных гармониках, на высоких производится интегральная оценка гармоник, но зато точно отслеживается динамика изменения их огибающей во времени – аналогично тому, как это происходит при оценке высоты тона.

В итоге, в низкочастотной области слух оценивает значение основной частоты фонации и ее первых обертонов, и по ним определяет высоту голоса; в верхней части слух точно оценивает изменение огибающей во времени, что позволяет ему выделить формантную картину, которая служит базовой информацией для верхних отделов мозга при определении фонетического значения отдельных фонем, слогов и др.

Таким образом, при анализе акустических параметров речевого сигнала в современных специализированных программах оцениваются следующие характеристики:

- уровнеграмма и все связанные с ней параметры (динамический диапазон, распределение мгновенных значений сигнала, текущая мощность и др.);
- одномерный спектр (распределение формантных областей);
- трехмерный спектр (изменение формы огибающей во времени);
- спектрограммы (широкополосные, узкополосные, слуховые), из которых могут быть получены такие характеристики, как изменение основной фонационной частоты во времени, изменение формантных областей, распределение гармоник голосового источника, временная структура импульсов звукового давления и др.

Кроме того, в ряде программ предусмотрена операция расчета нелинейной маскировки составляющих речевого сигнала, удаление неслышимых компонент, расчет распределения формантных полос с учетом их ширины и добротности. др. Общая картина анализа речевого сигнала, обычно производимая в современных компьютерных программах, показана на рисунке 8.

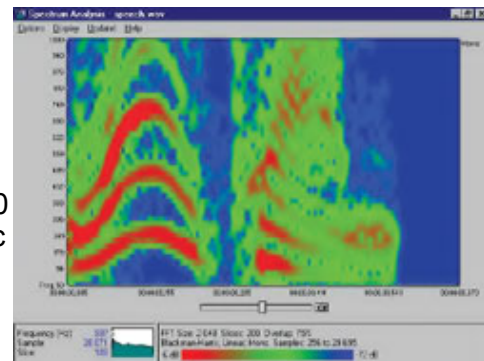


Рис. 6. Спектрограмма речевого сигнала

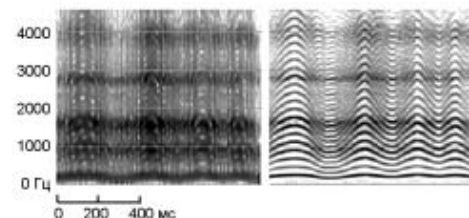


Рис. 7. а) Широкополосная спектрограмма; б) узкополосная спектрограмма

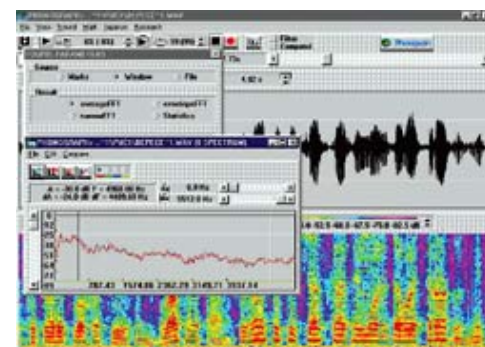


Рис. 8. Пример анализа речевого сигнала

Часть 17. Слух и речь. Часть 4. Субъективные и объективные методы оценки разборчивости речи

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 120° в горизонтальной и 90° 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 ревербирующем помещении и др. Шумы могут создаваться вентиляцией, внешними проникновениями, шумами аппаратуры, публикой, электронной аппаратурой и др.

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).

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 12 дБ/окт и с частотой среза не ниже 100 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.

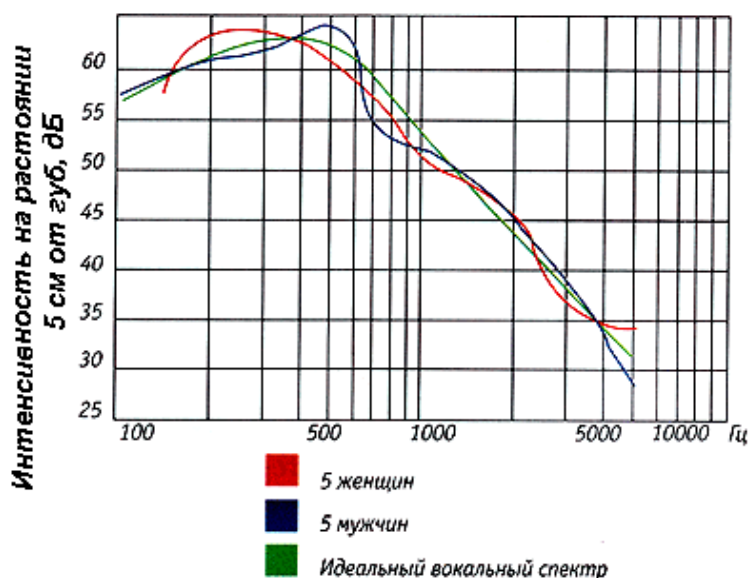


Рис. 1 Частотная зависимость спектральной плотности речи

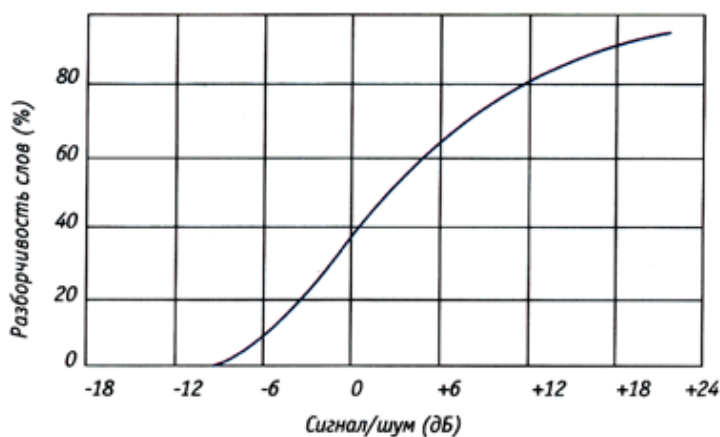


Рис. 2 Зависимость словесной разборчивости от отношения сигнал/шум для широкополосного маскирующего шума

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 падает 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.

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 (частота 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.

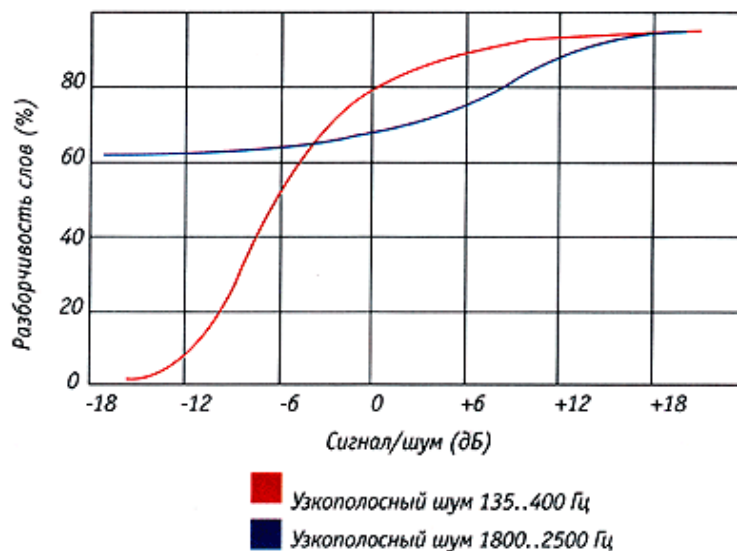


Рис. 3 Зависимость словесной разборчивости от отношения сигнал/шум для н/ч и в/ч узкополосного шума

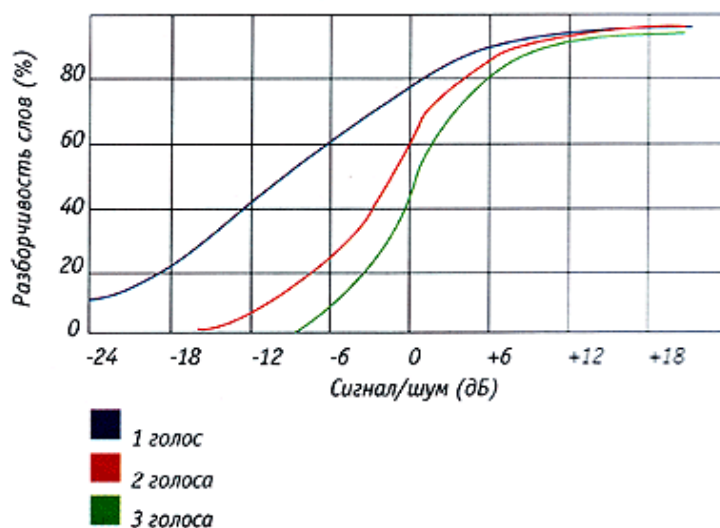


Рис. 4 Зависимость словесной разборчивости от отношения сигнал/шум при воздействии других голосов

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.

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 статистической 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

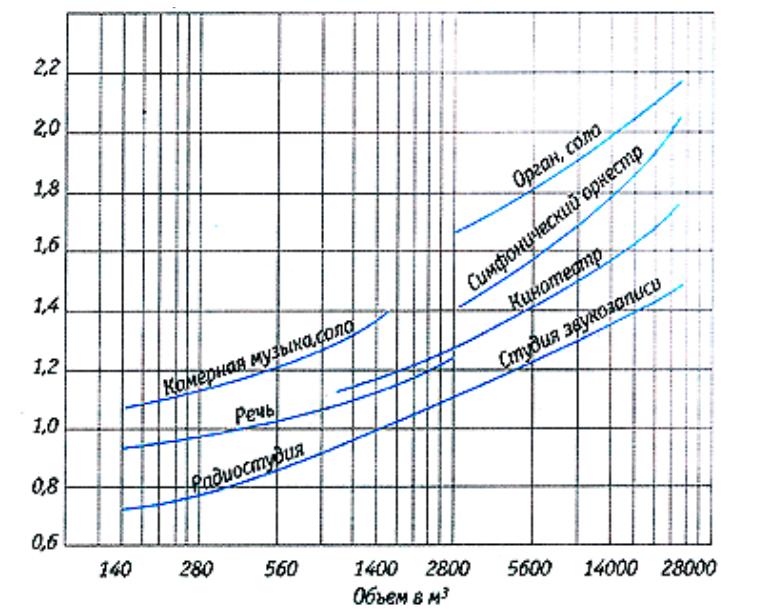


Рис. 5 Оптимальное время реверберации для помещений разного объема

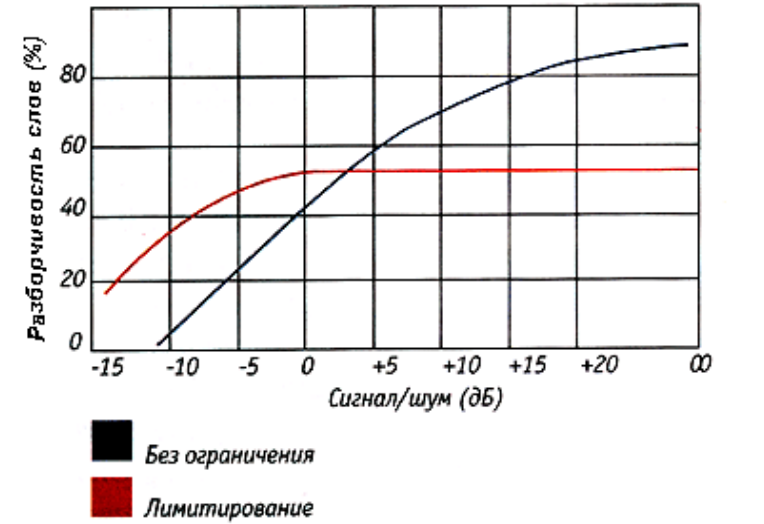


Рис. 6 Влияние клиппирования сигнала на процент словесной разборчивости речи

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 циклическая 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.

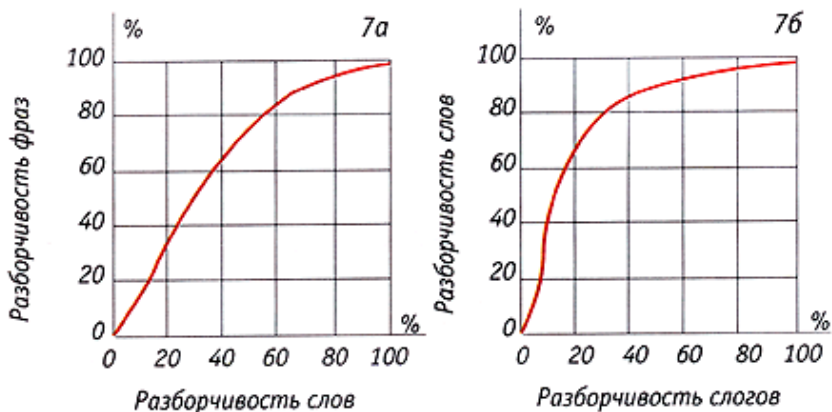


Рис. 7а - связь фразовой и словесной разборчивости;
7б - связь словесной и слоговой разборчивости

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.

Часть 17. Слух и речь Часть 4. Субъективные и объективные методы оценки разборчивости речи, часть 2

Как уже было сказано в предыдущей части данной статьи (5/2002), разборчивость является некоторой интегральной оценкой речевого сигнала и, в соответствии с международным стандартом ISO/TR 4870, определяется как *"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.

Для определения разборчивости используются на практике, в частности, при оценке акустических свойств аудиторий, театральных, концертных залов, студий и др., а также при разработке и установке различных систем звукоусиления, методы (стандарты ГОСТ 25902-83, ГОСТ 51061-97, ANSI S3.2 и др.), о которых уже было рассказано в предыдущей части статьи.

Однако, учитывая, что процесс организации субъективных экспертиз достаточно сложная, длительная и дорогостоящая процедура, на протяжении уже многих лет ведутся работы по поиску объективных методов оценки разборчивости, позволяющих получить быстрые и автоматизированные оценки, хорошо совпадающие с субъективными экспертизами.

В настоящее время разработано достаточно большое количество *объективных методов*, основные из которых: **AI** (articulation Index) - индекс артикуляции; **%AL_{cons}** (percentage Articulation Loss of Consonants) - процент артикуляционных потерь согласных; **STI** (speech transmission index) - индекс передачи речи; **RASTI** (rapid speech transmission index) - быстрый индекс передачи речи; **SII** (speech intelligibility index) - индекс разборчивости речи и др. (стандарты ISO/TR-4870, ANSI S3.2, S3.5; IEC 268-16 и др.).

На их анализе остановимся более подробно в данной статье, поскольку применение современных методов компьютерной оценки разборчивости речи является очень полезным при работе звукорежиссера в студиях и театрально-концертных залах (особенно если в них установлена система звукоусиления), а также при оценке качества речевых сигналов при передаче по каналам радиовещания, телефонии, в системах перевода речей и пр.

Объективные методы оценки разборчивости речи

Многолетний опыт проектирования залов различного назначения (аудиторий, лекционных залов, кинозалов, театральных залов и др.) и результаты многочисленных исследований показали, что разборчивость речи в помещении определяют следующие акустические характеристики:

- уровень прямого речевого сигнала во всех точках зала;
- уровень внешних и внутренних шумов;
- время реверберации;
- структура, уровень и направление прихода отраженных сигналов.

При этом для хорошей разборчивости речи требуются: достаточно высокий уровень отношения полезного сигнала к уровню шума (для студий звукозаписи он должен соответствовать кривой NC-15) (рис.1), небольшое время реверберации (в пределах до 1,1 с) и структура отражений, характеризующаяся отсутствием эха и наличием интенсивных ранних отражений при низком уровне поздних.

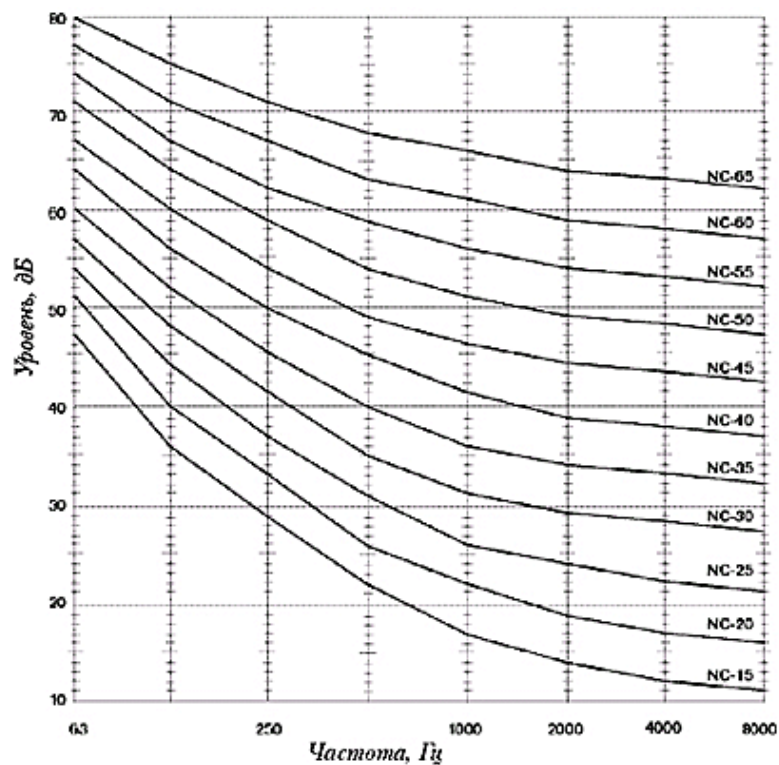


Рис. 1 Стандартизированные кривые уровня шумов NC

Влияние реверберационного процесса на структуру речевого сигнала можно отчетливо увидеть на примере осциллограмм, записанных в заглушенной камере и в помещении с большим значением реверберации (рис.2). Естественно, что при таком существенном изменении временной структуры речевого сигнала процесс его распознавания существенно ухудшается.

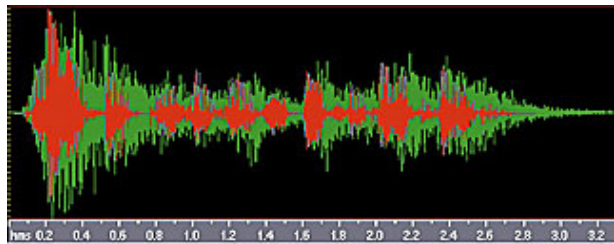


Рис. 2 Осциллограмма речевого сигнала в заглушенной камере и в помещении

Существующие в настоящее время методы объективной оценки разборчивости в той или иной степени связаны с перечисленными акустическими характеристиками. Кроме того, на разборчивость речи влияют параметры системы звукоусиления, если она установлена в зале (рис.3), о чем уже было сказано в первой части статьи.

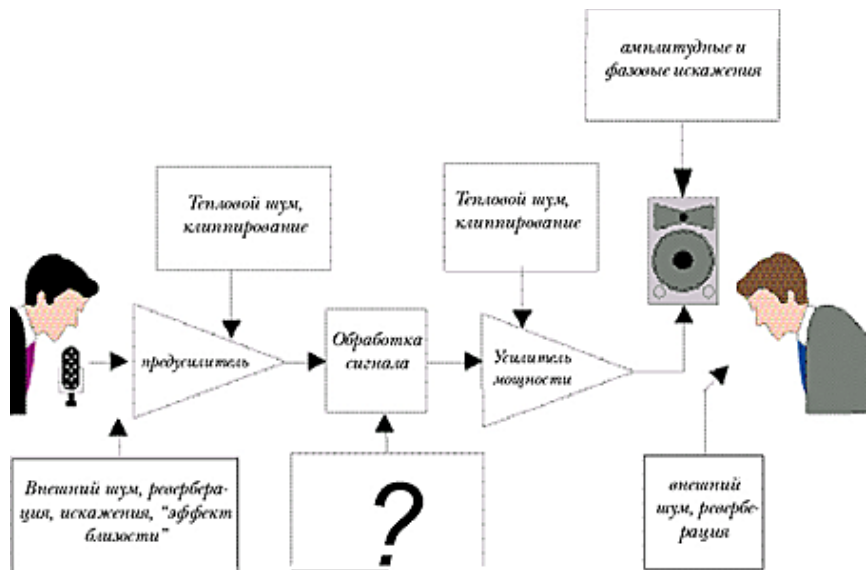


Рис. 3 Параметры системы звукоусиления, влияющие на разборчивость

Объективные методы оценки разборчивости можно разделить на две группы.

Первая аддитивная группа базируется на том предположении, что результирующая разборчивость речи определяется суммой вкладов в отдельных частотных полосах, а величина вклада зависит от отношения сигнала к шуму в каждой полосе.

К этой группе относится несколько методов.

1. **AI** - Индекс артикуляции.

Использование индекса артикуляции было предложено Г. Флетчером в лаборатории Bell Telephone Laboratories (США) в 1940 году.

Идея заключается в том, что весь частотный диапазон речевого сигнала разделяется на некоторое количество полос (обычно на 20), в пределах которых определяется отношение сигнала к шуму. Предполагается, что каждая частотная полоса вносит свой независимый вклад в разборчивость речи. Ширина полос выбрана таким образом, чтобы вклад каждой полосы в результирующую разборчивость был одинаковым.

В каждой полосе рассчитывается отношение "сигнал/шум", взвешивается и суммируется для получения индекса артикуляции: $AI = \sum AI_j$, где AI - индекс артикуляции, $0 < AI < 1$; AI_j - вклад в индекс в j -той полосе частот, $0 < AI_j < 0,05$. Причем, для отношения "сигнал/шум" больше 30 дБ значение AI_j принимается равным 0,05, а для отношения "сигнал/шум" меньше 30 дБ эти значения определяются как: $AI_j = 0,05(S/N)_j/30$. Отношение сигнала к шуму в отдельных полосах частот определяется по пиковым значениям речевого сигнала и по эффективным величинам шума. Причем, в шум включаются все мешающие передаче речи сигналы: шумовой фон помещения, реверберационная составляющая речевого сигнала и шум электроакустического тракта.

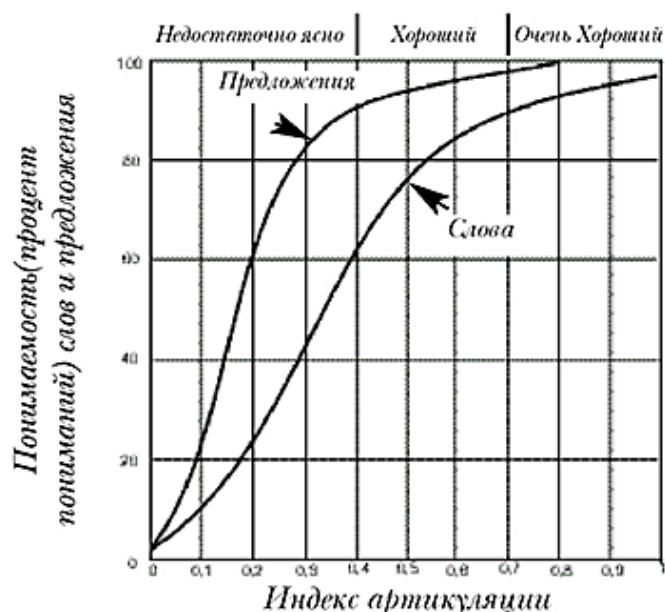


Рис. 4 Соотношение между разборчивостью и индексом артикуляции AI

Значения AI ниже 0,3 соответствуют плохой разборчивости; от 0,3 до 0,5 - удовлетворительной; 0,5...0,7 - хорошей; выше 0,7 - очень хорошей. Соотношение между артикуляционным индексом и словесной разборчивостью для английской речи показано на рисунке 4. Как следует из определения, основное влияние на величину индекса артикуляции оказывает уровень шумов в помещении, примеры распределения значений AI в разных точках аудитории при уровне шумов, соответствующих кривым NC-50 и NC-35, представленных на рис.1, показаны на рис.5а и 5б.

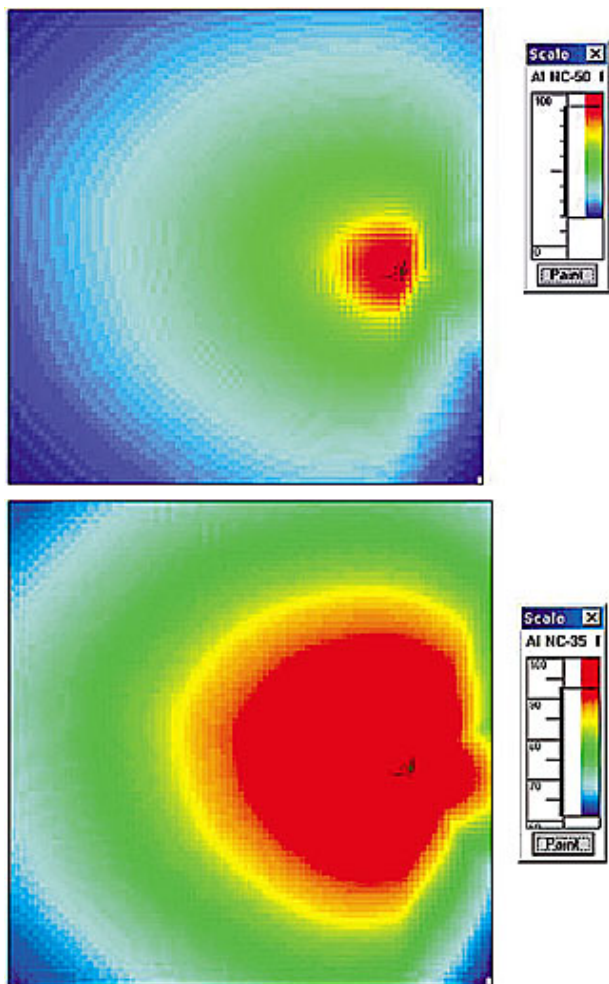


Рис. 5 Распределение индекса артикуляции AI в разных точках аудитории при разном уровне шумов: а-NC-50; б-NC-35

Видно, что в первом случае, значение индекса артикуляции падает от 0,93 (93%) вблизи оратора до 0,18 (18%) в дальней точке, то есть помещение имеет в основной зоне плохую разборчивость ниже 0,7 (70%). Во втором случае AI падает от 93% до 62%, в основной зоне помещения он превышает 70%, следовательно, только уровень шумов ниже NC-35 позволяет обеспечить приемлемую разборчивость речи в аудитории без звукоусиления.

Необходимо отметить, что в России этот метод был подробно разработан в трудах М.А. Сапожкова и др., он получил название "метода формантной разборчивости", при этом методика его определения приобрела некоторые изменения. Для расчетов выбирается частотная полоса речевого сигнала 150...7000 Гц, поскольку внутри этой полосы располагаются практически все форманты гласных и согласных русской речи. (Определение формант как частотных полос, где имеются максимумы акустической энергии, и их расположение в частотной области было рассмотрено в предыдущих частях этой статьи). Эта полоса делится на 20 подполос, вероятность появления формант в каждой полосе равна 0,05.

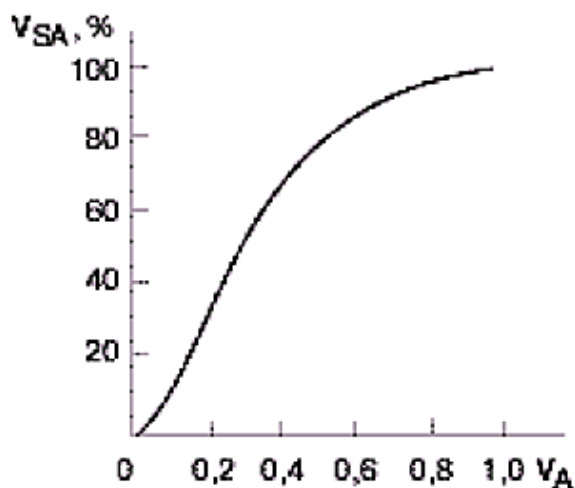


Рис. 6.а. Определение коэффициента K_{ϕ}

В результате маскировки шумом уменьшается вероятность восприятия формант (в первую очередь для согласных звуков, потому что они ниже по уровню), а следовательно и уровень разборчивости речи. Общий коэффициент формантной разборчивости также определяется как $AI = \sum AI_j$, где $AI_j = 0,05K_{\phi}$. Коэффициент K_{ϕ} называется коэффициентом восприятия формант (или коэффициентом разборчивости в к-полосе). Он определяется следующим образом: рассчитывается разница E_{ϕ} между средним уровнем речи в каждой полосе B_p и средним уровнем шумов и помех в этой же полосе $B_{ш}$: $E_{\phi} = B_p - B_{ш}$. Из полученного значения E_{ϕ} в каждой полосе определяется по графику значение коэффициента K_{ϕ} (рис.6а). Отсюда по вышеуказанной формуле для каждой полосы рассчитывается коэффициент AI_j , и затем суммарный коэффициент формантной разборчивости (индекс артикуляции). Связь этого коэффициента с величиной слоговой разборчивости для русской речи показана на рис.6б. Интересно, что при потере почти половины формантной разборчивости слоговая разборчивость составляет 80%, то есть речь обладает достаточным запасом по информационной избыточности.

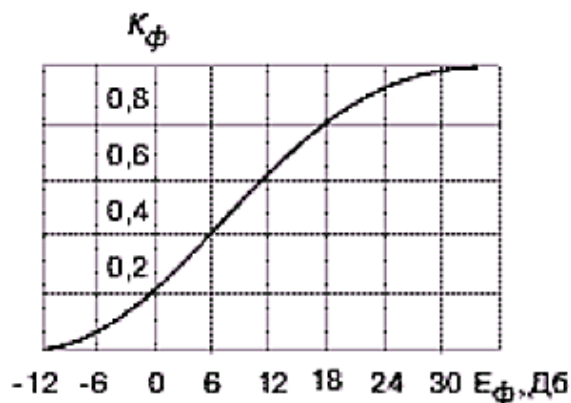


Рис. 6.б. Соотношение между слоговой разборчивостью и индексом артикуляции AI

(русская речь)

Несмотря на простоту применения, метод расчета артикуляционного индекса имеет ряд существенных недостатков: он учитывает только влияние шума на разборчивость, и не учитывает влияния других факторов (времени реверберации, структуры отражений и др.). Поэтому он больше подходит для оценки разборчивости в системах связи, а при оценке разборчивости в помещении с временем реверберации $T_{рев} > 0,5$ с он дает не очень хорошее совпадение с субъективными оценками. Этот метод считается устаревшим, и в современной практике используются в основном другие критерии.

2. SII - индекс разборчивости речи. Он был предложен как дальнейшее развитие метода AI и даже включен в американский стандарт для оценки разборчивости речи ANSI S3.5-1997.

В стандарте предлагается четыре измерительные процедуры, каждая использует различное число и размер частотных полос в пределах диапазона 150...8500 Гц:

- критические полосы (21 полоса);
- третьоктавные полосы (18 полос);
- равные по вкладу критические полосы (17 полос);
- октавные полосы (6 полос).

В каждой из них вычисляется отношение "сигнал/шум" и затем рассчитывается суммарный коэффициент SII, который может изменяться в пределах от 0 до 1.

Поскольку здесь используются критические полосы (о них было написано в предыдущих статьях по психоакустике), метод дает несколько лучшую корреляцию с субъективными экспертизами. Однако, проблемы учета искажений во временной области (реверберация, эхо и др.) по-прежнему остаются.

3. %AL_{cons} - процент артикуляционных потерь согласных (percentage Articulation Loss of Consonants).

Этот метод был разработан В. Пьютцем из Голландии - его первая статья была опубликована в 1971 году в JAES. С того времени он опубликовал много работ по данному направлению, в 1984 году разработал алгоритм для измерений с помощью TEF-анализатора. Идея заключается в том, что для объективной оценки разборчивости вычисляется процент потери согласных, которые в основном и определяют смысловое содержание речи. Чем больше этот процент, тем хуже разборчивость речи в помещении. Максимально допустимое значение - 10% (при условии, что помещение относительно свободно от маскирующего шума). Для помещений, используемых для целей обучения, и для систем оповещения он должен составлять не больше 5%.

Первая из предложенных Пьютцем формул для расчета процента потери согласных **ALcons** имела следующий вид:

$$\%AL_{\text{cons}} = 200D^2RT_{60}^2(N)/VQM, \text{ где:}$$

D - расстояние от источника звука (оратора, певца или громкоговорителя) до самого удаленного слушателя, м;

RT_{60} - стандартное время реверберации, с;

V - объем помещения, м^3 ;

Q - фактор направленности источника звука для расчетной точки;

M - поправочный коэффициент, учитывающий влияние различных акустических и электроакустических эффектов (как правило равен единице);

N - коэффициент, учитывающий использование в системе нескольких источников.

В случае, если используется несколько излучателей для озвучивания помещения и все они одинаковы, то N определяется как отношение общего количества акустических систем к числу систем, обеспечивающих прямой звук в расчетной точке.

Как следует из этой формулы, в ней учитывается влияние объема помещения, расстояние до источника звука, коэффициенты направленности источников звука, отношение прямой энергии к ревербирующей и т.д. Из нее также следует, что для достижения уровня $\%AL_{\text{cons}} = 10\%$ время реверберации не должно быть больше 1,1 с. Было также установлено, что повышение уровня давления, создаваемого громкоговорителями, приводит к улучшению разборчивости до тех пор, пока уровень сигнала не достигает значения +25 дБ по отношению к уровню фонового шума. Дальнейшее повышение уровня давления (излучаемой акустической мощности) вопреки обычным представлениям не сказывается на степени разборчивости, так как это приводит к росту уровня как прямого, так и диффузного звуков.

Этот метод широко используется на практике (особенно в США), он позволяет учесть влияние реверберации на разборчивость речи и дает хорошую корреляцию в ряде случаев с субъективными оценками.

Примеры расчета значений $\%AL_{\text{cons}}$ для разных точек помещения (речевой аудитории) при времени реверберации 3,5 и 1,2 с (рис.7а,б) показывают, что при большом времени реверберации значения $\%AL_{\text{cons}}$ достигают 26% процентов в дальних точках помещения, что недопустимо. В то же время в более заглушенной аудитории, на большей части площади $\%AL_{\text{cons}}$ не превышает 10%, что обеспечивает вполне удовлетворительную разборчивость речи.

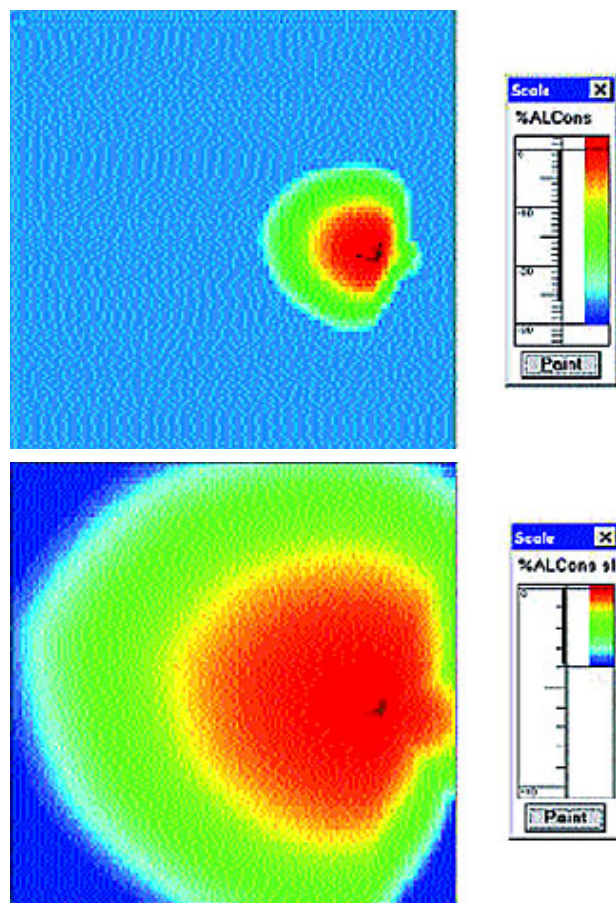


Рис. 7 Распределение в помещении

значений $\%AL_{cons}$: а-Т=3,5; б-Т=1,2 с

Но у этого метода есть свои недостатки: он основан на измерении в одной третьоктавной полосе вокруг 2000 Гц, поскольку здесь сосредоточена основная энергия согласных звуков. Все другие частоты обычно не используются, что может приводить к значительным погрешностям, в частности, за счет частотной зависимости коэффициента направленности источника звука от частоты и соответствующего изменения по частотному диапазону отношения уровня прямого звука к ревербирующему. Кроме того, этот метод не учитывает целый ряд важных для разборчивости параметров: отношение "сигнал/шум", поздние отражения и эхо, спектр фонового шума и др. Правда, позднее Пьютцем была предложена новая методика, позволяющая учесть фоновый шум. При этом предлагается значения $\%AL_{cons}$, вычисленные по вышеприведенной формуле, умножить на коэффициент, учитывающий уровень шумов в помещении:

$$\%AL'_{cons} = \%AL_{cons} (1,071T^{-0,0285})^{25+(L_n - L_s)}, \text{ где:}$$

L_n - уровень шума в помещении, дБ; L_s - уровень речевого сигнала, дБ.

Однако проблемы, связанные с применением этого критерия (возможность эффективного применения в основном в помещениях сравнительно небольшого объема, неучитывание ранних отражений, эхо и др.) остаются, поэтому работы по усовершенствованию этого критерия все время продолжают.

Вторая группа методов для объективной оценки разборчивости построена на оценке модуляционной передаточной функции системы. К их числу относятся методы STI, RASTI, STIPA и др.

1. **STI** - Индекс передачи речи (*speech transmission index*).

Особая группа методов, позволяющая объективно оценить разборчивость речи в помещении, была разработана на основе использования модуляционной передаточной функции MTF (The Modulation Transfer Function).

Идея методов заключается в том, что, как было рассмотрено в предыдущих частях этой статьи, речевой сигнал представляет собой свертку импульсной характеристики звукового сигнала источника (для гласных звуков это импульсный сигнал с почти гармоническим спектром, для согласных это шумовой сигнал различной формы) с импульсной характеристикой голосового тракта. Как показывает анализ рис.8, речевой сигнал можно приблизительно рассматривать как некоторый широкополосный сигнал (с полосой 125...8000 Гц), модулированный другим сигналом с низкой частотой. Частота модуляции определяется скоростью, с которой человек произносит форманты (скоростью артикуляции). Эксперименты показали, что частоты модуляции в обычной речи находятся в диапазоне от 0,63 до 16 Гц, причем наиболее вероятные частоты модуляции находятся в области 5...7 Гц.

Уменьшение глубины модуляции делает речевой сигнал более похожим на шумовой и, следовательно, приводит к уменьшению его разборчивости. При прохождении речевого сигнала через любой тракт передачи или помещение за счет реверберации, искажений и шумов происходит уменьшение разборчивости сигнала, что может быть оценено по степени уменьшения его глубины модуляции (изменения формы огибающей сигнала).



Рис. 8 Осциллограмма речевого сигнала

Поэтому в работах Хоутгаса и Стинекена в 1971 году был предложен следующий метод оценки разборчивости: весь речевой диапазон от 125 Гц до 8000 Гц разбивается на семь октавных полос. На вход испытываемой системы, например, на акустическую систему, установленную в точке помещения, где должен находиться диктор (оратор), подается октавный шумовой сигнал с центральными частотами 125; 250; 500 Гц; 1; 2; 4 и 8 кГц, распределение интенсивности которого по октавным полосам совпадает с распределением интенсивности речевого сигнала. Сигнал этот модулируется гармоническим сигналом со стопроцентным индексом модуляции (рис.9): $I(t) = I_0(1 + \cos 2\pi Ft)$. При этом характеристика направленности акустической системы, через которую сигнал излучается в помещение, должна быть близка к характеристике направленности ротового отверстия. Частоты модулирующего сигнала меняются в пределах от 0,63 до 12,5 Гц с третьоктавным интервалом (всего четырнадцать частот). В точке, где должен располагаться слушатель, сигнал принимается ненаправленным микрофоном, при этом за счет шумов и процессов реверберации в помещении форма сигнала изменяется, и глубина его модуляции уменьшается.

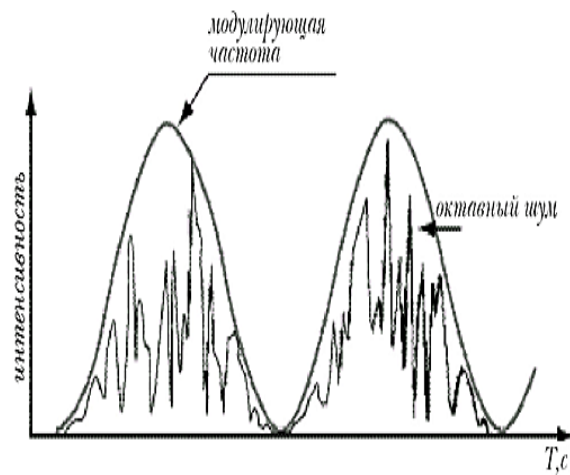


Рис. 9 Вид модулированного сигнала

Сигнал теперь может быть представлен в форме: $I(t) = I_0(1 + m \cos(2\pi Ft + \phi))$, где:

m - коэффициент, характеризующий уменьшение глубины модуляции. Он является функцией частоты модуляции F и зависит от времени реверберации в помещении и отношения "сигнал/шум" (зависимость величины $20 \lg(m(F))$ от частоты F называется модуляционной передаточной функцией);

I_0 - амплитудное значение сигнала;

F - частота сигнала;

t - время;

ϕ - начальная фаза.

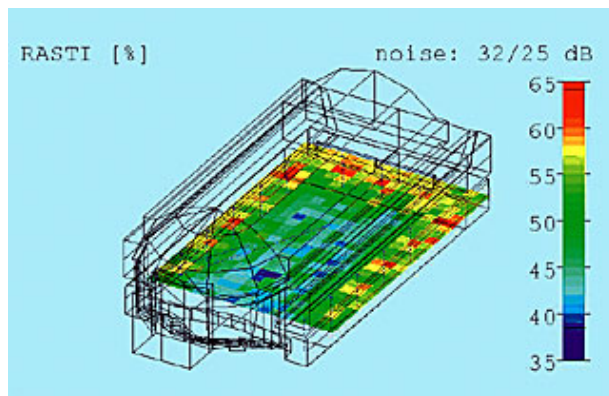


Рис. 10 Шкала CIS

В предлагаемом методе измерения проводятся для 98 точек для семи октавных полос подаваемого шума и 14 значений частот модуляции. Полученная матрица из 98 точек используется для расчета STI (speech transmission index) путем взвешивания среднего значения $m(F)$ в каждой полосе частот в соответствии с их общим вкладом в разборчивость. Пример такой матрицы, полученной с помощью компьютерной станции MLSSA, приведен ниже на

Методы измерения разборчивости с помощью STI введены в международный стандарт IEC 268-16. Этот метод предполагает проведение измерений в широком диапазоне частот и поэтому учитывает частотную зависимость времени реверберации, неравномерность АЧХ и другие частотно-зависимые эффекты, что в результате дает достаточно хорошую корреляцию с субъективными оценками.

Для того, чтобы была возможность сравнить результаты оценок разборчивости, полученных различными методами, была разработана общая шкала разборчивости CIS (стандарт IEC-60849), позволяющая пересчитывать значения разборчивости, полученной разными методами (как субъективными, так и объективными) одна в другую (рис.10). Например, значение STI равное 0,7, соответствует значению $\%AL_{cons}$ 3,8%, что соответствует 94% слоговой разборчивости и обеспечивает хорошую разборчивость в помещении. Пределы изменения STI для обеспечения хорошей разборчивости составляют 0,6...1,0.

2. RASTI - быстрый индекс передачи речи (rapid speech transmission index).

Поскольку расчет разборчивости вышеуказанным методом представляет собой довольно трудоемкую процедуру, фирмой Bruel & Kjaer был разработан альтернативный упрощенный метод, предусматривающий измерения только в двух октавных полосах 500 Гц при частотах модуляции 1,02/2,03/4,07/8,14 Гц, а также 2 кГц при частотах модуляции 0,73/1,45/2,90/5,81/11,63 Гц. Дальше процедура измерений происходит как и в предыдущем случае, строится матрица, рассчитывается усредненное значение модуляционной передаточной

функции в полосах частот и по нему определяется нормализованный индекс RASTI.

Для реализации этого метода фирмой Bruel & Kjaer были разработаны соответствующая аппаратура (B&K3361 RASTI Meter) и программное обеспечение, поэтому эта методика принята практически во всех европейских стандартах. Однако поскольку измерения проводятся только в двух октавных полосах, то они дают несколько завышенные значения, поскольку не учитывают неравномерность АЧХ и ФЧХ в системе. С целью еще большего ускорения процедуры измерений фирма предложила метод, при котором сигналы предъявляются одновременно в двух октавных полосах со всеми модулирующими сигналами одновременно.

Измерения производятся в непрерывном режиме и позволяют производить параллельный анализ в двух октавных полосах. Интервал времени для анализа требуется не менее 8 с. Для хорошей разборчивости требуется, чтобы значения RASTI в разных точках помещения были не ниже 0,6. Для его оценки измерения проводятся в разных точках помещения и строятся кривые его распределения. Пример показан на рис.11.

MTF Matrix (Calibrated)

Frequency-Hz	125	250	500	1000	2000	4000	8000
level dB	-24.5	-22.9	-23.2	-24.8	-26.4	-32.2	-65.5
m-correction	1.000	1.000	1.000	1.000	1.000	0.999	0.605
0.63	0.988	0.987	0.977	0.936	0.791	0.552	0.071
0.80	0.982	0.984	0.975	0.933	0.794	0.542	0.067
1.00	0.982	0.984	0.975	0.933	0.794	0.542	0.067
1.25	0.978	0.982	0.974	0.935	0.793	0.553	0.060
1.60	0.978	0.982	0.974	0.935	0.793	0.553	0.060
2.00	0.973	0.979	0.973	0.936	0.798	0.548	0.068
2.50	0.967	0.976	0.972	0.933	0.793	0.545	0.052
3.15	0.958	0.971	0.968	0.930	0.791	0.558	0.058
4.00	0.946	0.967	0.966	0.929	0.787	0.535	0.047
5.00	0.935	0.961	0.962	0.929	0.793	0.536	0.060
6.30	0.920	0.952	0.958	0.925	0.787	0.544	0.072
8.00	0.905	0.946	0.953	0.923	0.786	0.555	0.074
10.00	0.880	0.933	0.947	0.920	0.786	0.549	0.059
12.50	0.854	0.920	0.940	0.918	0.786	0.551	0.057
octave TI	0.926	0.967	0.975	0.874	0.693	0.528	0.109

STI value= 0.704 (0.751 modified) ALcons= 3.8% Rating= GOOD

ESC to exit or F1 to print.

MLSSA: STI

Рис. 11 Пример распределения RASTI значений в помещении

Следует отметить, что, строго говоря, методы оценки разборчивости с помощью модуляционных передаточных функций применимы только к линейным системам. Большинство же реальных систем нелинейны, особенно если речь идет об оценке разборчивости в системах коммуникаций (в телефонных каналах, например). Воздушное пространство в помещении можно считать линейной средой, но надо строго контролировать линейность измерительной аппаратуры.

Поиски простых и надежных методов для оценки разборчивости все время продолжаются. В 2001 году фирмой Bose была предложена методика измерений STI со специальным оптимизированным тестовым сигналом, представляющим собой модулированный шум в шести октавных полосах с двумя одновременными частотами модуляции в каждой октаве. Эта методика получила название STI-PA (то есть предназначенная для оценки систем звукоусиления PA).

Кроме того, развитие цифровых методов обработки звука и увеличение скорости компьютерной обработки позволило создать методику (она была предложена Шредером), с помощью которой значения модуляционной передаточной функции можно рассчитать из импульсных измерений сигнала с его соответствующей фильтрацией. Современные компьютерные метрологические станции TEF, MLSSA и др. используют именно эти методы для расчета разборчивости.

Необходимо отметить, что, кроме наиболее распространенных сейчас методов оценки разборчивости речи, использующих модуляционную передаточную функцию, достаточно широко используются при оценке помещений и другие методы, например, оценка разборчивости речи по коэффициенту четкости (ясности).

C50 - коэффициент четкости. Коэффициент C50 определяет четкость (ясность) звучания музыки, и вычисляется как отношение ранней к поздней энергии отраженных звуков в помещении. В 1996 году Маршалл предложил использовать этот коэффициент для оценки разборчивости речи, для этого необходимо измерять отношение Еран/Епозд в нескольких частотных полосах, со специальным взвешиванием для речевого сигнала. Измерения проводятся с помощью анализа импульсной характеристики с последующим компьютерной обработкой - к ранней относится энергия сигнала в пределах первых 50 мс. Несмотря на простоту измерений, метод дает неплохую корреляцию с субъективными оценками, но он также обладает определенными недостатками: учитывает только влияние реверберационного процесса, слабо учитывает структуру мешающей части отклика, в которой могут быть поздние отражения большой интенсивности.

Эти отражения часто приводят к появлению эха, снижающему разборчивость речи. Указанные недостатки в значительной мере ликвидированы в критерии, предложенном Низе, и названным "степенью эха". Низе ввел новую 33 мс границу между полезной и мешающей энергией, причем полезная энергия берется с весовой функцией, смягчающей резкую границу деления импульсного отклика. Мешающей же считается не вся

энергия, а только энергия отражений, превышающих экспоненциальную кривую процесса реверберации данного помещения. Известны и другие модификации этого метода.

В настоящее время созданием международных рекомендаций по оценке разборчивости речи занимается специальная группа Комитета Стандартов AES (председатель П. Мапп). Специалисты этой группы занимаются сравнительным анализом различных методов оценки разборчивости. В процессе этих работ были получены, в частности, достаточно простые аналитические соотношения между разными коэффициентами по оценке разборчивости речи, которые могут быть полезны в практической работе по оценке помещений.

В заключение следует отметить, что проблема оценка разборчивости речи чрезвычайно сложна, поскольку она зависит не только от факторов, которые поддаются объективным измерениям (время реверберации, уровень шумов, структура ранних отражений и др.), но и от таких субъективных факторов, как степень владения языком, дикция оратора, степень знакомства слушателя с содержанием сообщения, состоянием его слуха и др. Поэтому создание автоматизированной компьютерной системы объективной оценки разборчивости речи - проблема трудная, но, поскольку она чрезвычайно важна для развития коммуникационных систем, и к ее решению привлечены значительные силы специалистов в различных странах, можно надеяться, что такая достаточно надежная система будет создана.

Часть 17. Слух и речь Часть 5 Акустические характеристики вокальной речи

В предыдущих статьях из серии "Слух и речь" были рассмотрены способы образования речи и ее акустические характеристики. В данной статье остановимся на специфических акустических характеристиках вокальной речи, работа с которой составляет значительную часть деятельности звукорежиссеров, особенно при известном уровне вокального искусства на современной эстраде.

Термин "вокальная речь" обозначает то же, что и термин "пение", но пение – это прежде всего музыкально-эстетический термин ("искусство пения"). Под термином "вокальная речь" понимается рассмотрение процесса пения с позиций акустики, поэтому при дальнейшем изложении будет использоваться в основном первый термин.

Информацию, которую несет речевой сигнал, принято делить на семантическую (смысловую, логическую) и эмоционально-эстетическую.

Если в разговорной речи основным является первый вид информации, то в вокальной речи главная задача состоит в передаче эмоционально-эстетической информации.

Эти задачи, а также особенности их реализации (оперное, концертное, хоровое пение, эстрадное и др.), привели к тому, что вокальная речь, несмотря на наличие общего механизма звукообразования, имеет ряд существенных отличий от обычной речи. Они заключаются, прежде всего, в следующем:

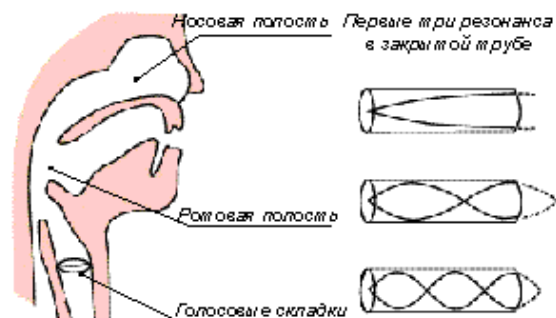
- специфике организации дыхания;
- использовании различных регистров;
- наличии специфической структуры формант, характеризующейся появлением особых певческих формант;
- способности к подстройке формант в верхнем диапазоне певческого голоса;
- наличии амплитудной и частотной модуляций (вibrато и тремоло);
- больших эмоциональных флюктуациях параметров звукообразования;
- особой системе обратной связи (эффект Томатиса) и др.

Все это обуславливает специфику акустических характеристик вокальной речи – более широкий частотный диапазон, большой динамический диапазон, особое спектральное распределение мощности и др., – и также создает специфические проблемы с разборчивостью.

Рис. 1 Общая структура вокального тракта

Процесс звукообразования вокальной речи (пения) включает в себя все основные этапы, которые используются при создании обычной речи: аспирации (генерации), фонации (для гласных и звонких согласных), артикуляции и излучения (рисунок 1). Однако перечисленные особенности вокальной речи вносят свои коррективы в организацию процессов, связанных с использованием как голосовых источников, так и резонаторных полостей (артикуляционного аппарата).

Начнем рассмотрение с резонаторных свойств, поскольку они создают наиболее заметные отличия в акустических характеристиках вокальной речи. К их числу прежде всего относятся появление певческой форманты, эффект



подстройки формант, отличия в первых формантах, особенности частотных диапазонов различных голосов и др.

Это обусловлено особыми требованиями к тембру певческого голоса, музыкальной экспрессивности, к мощности, обеспечивающей слышимость голоса на фоне аккомпанемента (например, симфонического оркестра), к способности воспроизводить широкий звуковысотный диапазон и др. Субъективные оценки тембра певческого голоса связаны, прежде всего, с его спектральными характеристиками (звонкость, яркость, полетность и др.), которые зависят как от спектрального состава звука, создаваемого голосовым источником, так и от использования резонансных свойств артикуляционных органов. Поэтому особые требования к тембру певческого голоса, существенно отличающиеся от требований к обычной речи, приводят в первую очередь к особенностям в настройке резонансов голосового тракта, т.е. его формантных частот.

Певческая форманта

Рис. 2 Сравнительный уровень звукового давления при пении и речи (Δ - изменение частоты фонации)

Для того, чтобы голос оперного певца (или концертного исполнителя) был слышен на фоне оркестра, необходимо, прежде всего, развивать значительно более высокий уровень звукового давления – проще говоря, петь громче. Сравнительный анализ уровней звукового давления на расстоянии 0,5 м при пении и речи для разных частот фонации (высоты тона) показан на рисунке 2. Однако одного увеличения громкости оказывается недостаточно для решения этой проблемы, поэтому в процессе многолетнего совершенствования техники пения, прежде всего оперного (*bel canto*), были отработаны особые приемы перестройки спектральных характеристик голоса, в частности создание так называемой певческой форманты.

Рис. 3 Расположение формант при речи и пении

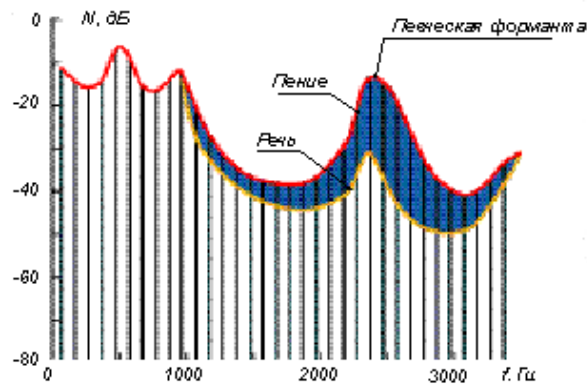
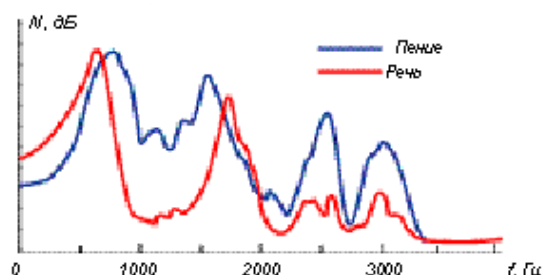
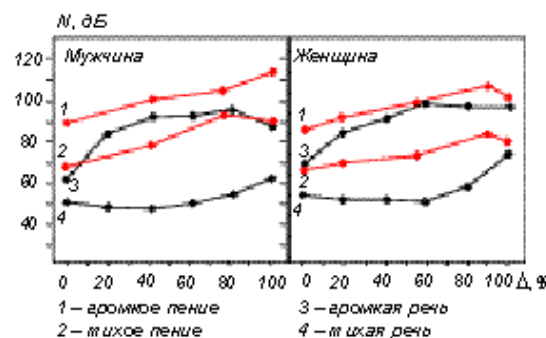
Если сравнить расположение формантных частот (т.е. резонансов вокального тракта) для одних и тех же гласных в обычной и вокальной речи, то отчетливо видно, что для всех гласных положение первой форманты на шкале частот мало изменилось, вторая форманта в случае вокальной речи сдвинута по частоте вниз, значительно изменились положения третьей, четвертой и пятой формант в сторону их совместного сближения (рисунок 3).

Рис. 4 Форма огибающего спектра при речи и пении

Если проанализировать спектральную огибающую любого из гласных звуков при речи и пении (рисунок 4), то отчетливо видно наличие выраженного пика в области 2...3 кГц. Этот пик создается в результате объединения (кластеризации) третьей, четвертой и пятой формант и носит название "певческой форманты".

Такая группа подчеркнутых формант наблюдается только в спектрах хорошо поставленных, чаще оперных, голосов, как показано в работах В.П.Морозова и Дж. Сандберга (наиболее известные ученые, занимающиеся проблемами акустики певческого голоса). При этом центральная частота располагается в зависимости от типа голоса в области: бас и баритон – 2,1...2,5 кГц, тенор – 2,5...2,8 кГц, сопрано – 3...3,5 кГц.

Для одного и того же певца эта форманта не смещается по частоте, она стабильно занимает одно и то же положение при пении нот разной высоты (рисунок 5а) или разных гласных (рисунок 5б). Таким образом, положение этой форманты не зависит ни от фундаментальной частоты (это свойство подстройки есть у других формант), ни от позиции других резонансов (нижних формант).



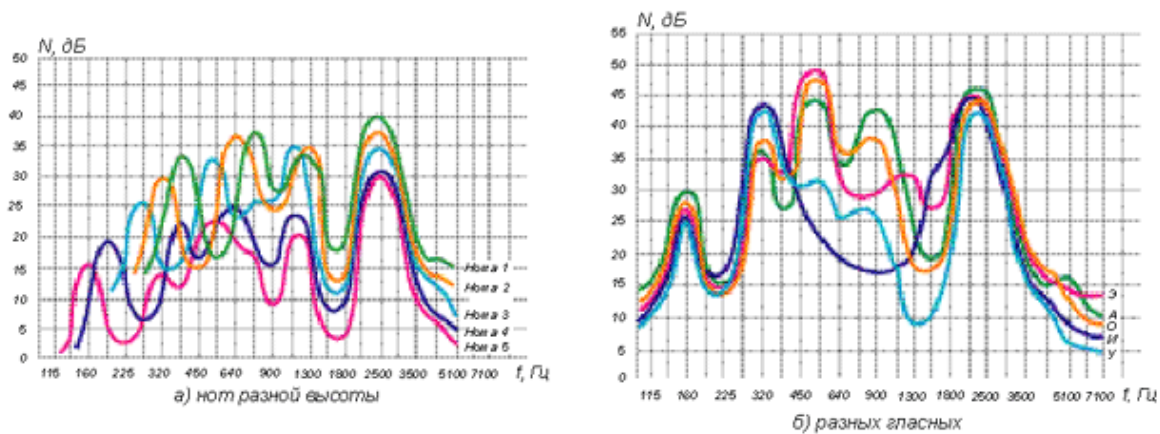


Рис. 5 Положение ВПФ при пении

Амплитуда певческой форманты зависит как от амплитуд объединяемых формант (третьей, четвертой и пятой), так и от амплитуд соответствующих обертонов в спектре голосового источника, а это, в свою очередь, зависит от особенностей использования голосового источника (в частности, скорости и силы сведения при колебаниях голосовых связок). Кроме того, поскольку при увеличении громкости огибающая спектра звука голосового источника меняется и увеличивается амплитуда высоких обертонов, то и амплитуда певческой форманты будет в громких звуках выше.

Исследования, выполненные различными учеными (результаты приведены, например, в замечательной книге В.П.Морозова "Биофизические основы вокальной речи"), показали, что при прослушивании записей певческих голосов, из которых электрическим путем вырезалась высокая певческая форманта, в голосе теряется звонкость, полетность, чистота и ясность тембра, голос звучит тускло, глухо, невыразительно.

Звучание высокой певческой форманты в изолированном виде (например, если отфильтровать в записи все частотные области, кроме третьего-четвертого резонансов) воспринимается как звонкий мелодичный звук типа соловьиной трели. Он слышен при пении любой гласной, усиливаясь при увеличении громкости. Эти опыты позволили установить, что от наличия и амплитуды высокой певческой форманты зависит звонкость певческого голоса. Интересно отметить, что у старинных скрипок (Страдивари, Амати, Гварнери), отличающихся особой звонкостью, полетностью звука, также в этой области частот (2,5...3 кГц) находится область выраженных резонансов (третья форманта). В работах В.П.Морозова была даже предложена количественная величина K – коэффициент звонкости голоса:

$$K = I_{\text{впф}} / I,$$

где: $I_{\text{впф}}$ – интенсивность спектра в области высокой певческой форманты, I – общая интенсивность голоса.

Измерения, проведенные у разных категорий певцов, показали, что у выдающихся певцов этот коэффициент достигает 33...35%, у неквалифицированных певцов – 5...11%, а у детей – 3%.

Однако основной причиной использования певческой форманты в оперных голосах является, по-видимому, повышение помехоустойчивости голоса.

Однако основной причиной использования певческой форманты в оперных голосах является, по-видимому, повышение помехоустойчивости голоса.

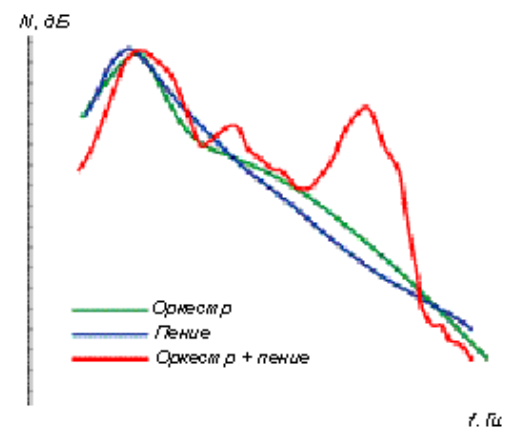
Средний уровень оркестра при достаточно громком исполнении в концертном зале составляет 90...100 дБ. Создание такого среднего уровня выше возможностей человеческого голоса. Кроме того, маскирующее действие оркестра зависит от частотного распределения спектральных уровней.

Рис. 6 Огибающая звуков оркестра, речи и пения

Если сравнить огибающую долговременного спектра для звуков оркестра и для звуков нормальной речи и пения (рисунок 6), то видно, что максимум энергии в звуках оркестра находится в области 400...500 Гц, затем ее уровень быстро падает и в области 2,5...3 кГц становится ниже на 20 дБ.

Примерно такое же распределение энергии и в звуках обычной речи (только ниже по уровню громкости), и поэтому она будет полностью замаскирована звуками оркестра. В певческом голосе до 30% энергии может быть сосредоточено в певческой форманте в области 2...3 кГц, которая к тому же совпадает с областью максимальной чувствительности слуха. Такая перегруппировка формант и позволяет услышать певческий голос на фоне оркестра. Это свойство голоса "перекрывать" оркестр называется "полетностью".

Следует отметить две слуховые особенности. Тихий звук ниже порога маскира (маскирующего звука) может быть услышан, если он вступает чуть раньше по времени или регулярно варьируется по времени. Звук



оркестра непрерывно меняется по уровню, поэтому необязательно все время слышать певческую форманту, достаточно слышать ее в определенные моменты, когда звук оркестра становится тише в этой области. Если маскер регулярно прерывать, то тихий звук можно не только услышать, но он может быть услышан как непрерывный во времени – слуховая система сама достраивает звучание, делает интерполяцию во времени.

Следующее преимущество певческой форманты относится к условиям излучения. Низкие частоты излучаются во всех направлениях (здесь рот – это ненаправленный источник звука), а высокие концентрируются только вперед, поэтому высокие частоты меньше поглощаются обратной стороной сцены, декорациями и др. Если в голосе есть певческая форманта, т.е. высокий уровень обертонов в высокочастотной части диапазона, то они попадают к слушателю менее поглощенными, чем низкочастотные компоненты, что помогает слушателю выделить голос певца из сопровождения оркестра (если он стоит лицом к слушателю). Это соответствует требованиям педагогов "сфокусировать" голос для слышимости в большой аудитории, что можно сделать только при наличии певческой форманты.

Механизм образования певческой форманты служит предметом многочисленных исследований. Результаты рентгенологических исследований показывают, что при пении профессиональных опытных певцов гортань занимает значительно более низкое положение, чем при речи (снижение до 15 мм). При этом удлиняется общая длина вокального тракта и глотки ("Звукорежиссер", 1/2002), что приводит к некоторому снижению второй форманты. При опускании гортани значительно расширяются боковые стенки дна глотки. При этом, если диаметр трубы гортани (в глубине которой на расстоянии примерно 20 мм от края находятся голосовые связки) составляет 1/6 от диаметра расширенной трубы глотки, то четвертая форманта снижается от 3,5 кГц (где она обычно находится у взрослого мужчины) до 2,8 кГц. Таким образом, за кластеризацию (объединение) высших формант в единую певческую форманту ответственна резонансная полость, которая образуется трубой гортани и дном расширенной глотки, слегка прикрытой надгортанником. Длина этой полости составляет примерно 1/6 длины всего тракта, нижний край полости ограничен голосовой щелью, верхний – кольцом надгортанника.

Такой метод образования певческой форманты за счет понижения позиции гортани и значительного расширения глотки используется не во всех манерах пения. Например, в китайской опере (или в некоторых видах средневекового пения) имеет место значительная певческая форманта, но без понижения позиции гортани, а, очевидно, за счет другой организации тракта.

Имеются и другие теории образования высокой певческой форманты: например, как краевых тонов, возникающих за счет турбулентного потока при прохождении струи воздуха через щель голосовых связок (как у флейты) и др.

Имеется значительное противоречие между использованием положения гортани в речи и пении: если при повышении высоты тона в обычной речи гортань занимает все более и более высокое положение, то при профессиональном мужском пении при повышении высоты тона положение гортани снижается, что требует специального обучения. Расширение гортани и снижение глотки ("открытое горло") при пении в грудном регистре являются признаками хорошо поставленного профессионального голоса.

Таким образом, спектральная перестройка голоса позволяет получить значительный энергетический выигрыш, повысить помехоустойчивость и звонкость голоса без увеличения энергетических затрат голосового источника. Не удивительно, что этот прием, найденный в процессе совершенствования техники пения, особенно пения "бель канто", получил такое широкое распространение.

Существуют виды пения, где певческая форманта не используется, например, хоровое пение, пение в бытовой манере, современное эстрадное пение, где баланс и тембр регулируются системой звукоусиления, и певческая форманта, т.е. усиление особой группы обертонов, может быть сформирована звукорежиссером. Вообще, качество пения у многих современных певцов – это почти исключительно искусство звукорежиссера, а не исполнителя, поэтому таким успехом и пользуется пение под фонограмму (под "фанеру").

Местоположение певческой форманты играет большую роль в классификации голосов, так как каждый голос – бас, баритон, тенор – имеет четкую позицию форманты, что, очевидно, отражает различия в длине вокального тракта и форме глотки. Следует еще отметить, что певческая форманта служит тембральной "униформой" для всех певческих гласных – в речи уровень третьей форманты отличается, например, на 28 дБ для "и" и "у", а уровень певческой форманты для всех гласных примерно одинаков.

Подстройка формантных частот

Следующей характерной особенностью акустических свойств вокальной речи является способность к подстройке формант. При пении на высоких тонах, особенно в женском пении, перестройка формант, приводящая к появлению певческой форманты, часто не используется. Во-первых, у высоких женских голосов другая структура вокального тракта (короче длина, трудно реализовать опускание гортани и расширение глотки и др.), во-вторых, наличие значительного подъема в области 3 кГц может привести к "пестроте" голоса, потому что при высокой фонационной частоте, например 700 Гц (F5), обертоны находятся на расстоянии 700 Гц друг от друга (редко), и поэтому отдельные обертоны будут то попадать, то не попадать в формантную область при сдвиге основной частоты, в отличие от обычной картины заполнения формантных областей при низкой фонационной частоте (рисунок 7). Поэтому в женском пении (и в высоких мужских голосах)

используется другой способ повышения помехоустойчивости голоса, обеспечивающий его слышимость на фоне оркестра, который получил название "подстройка формант".

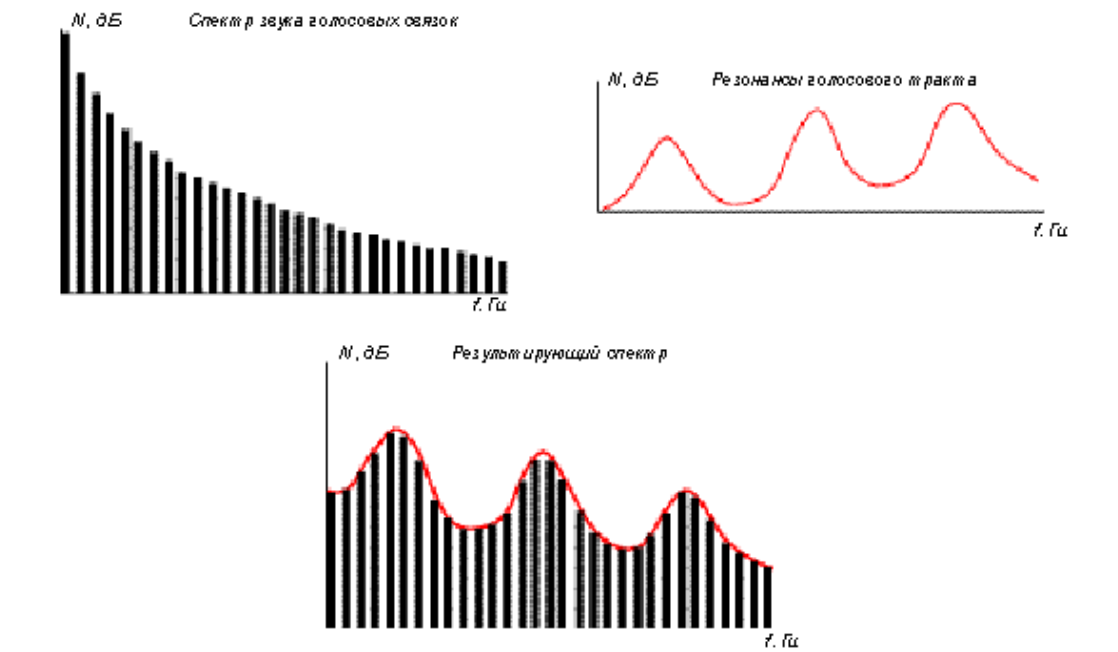


Рис. 7 Заполнение форматных областей обертонами

Большинство певцов использует в пении значение фундаментальной частоты выше, чем в обычной речи. В речи средняя фундаментальная частота составляет 110 Гц для мужского голоса и 200 Гц для женского (максимальный предел 200 и 350 Гц соответственно). В то же время в пении для высших нот сопрано, альты, тенора, баритона и баса значения фундаментальной частоты равны 1400 (C6), 700 (F5), 523 (C5), 390 (G4) и 350 (F4) Гц. Поэтому в речи первая форманта обычно выше, чем фундаментальная частота, а в пении во многих случаях первая форманта оказывается ниже ее. То есть, способность вокального тракта усиливать звук будет проявляться на частоте, где никакого звука от голосового источника не поступает.

Рис. 8 Связь между шириной открытия рта и частотой фонации

Как только первая форманта становится ниже по частоте, чем фундаментальная частота, певцы стараются подтянуть первую форманту к фундаментальной частоте, широко открывая рот. Это объясняет, почему открытие рта у певцов зависит от высоты тона, а не от типа гласной, как в обычной речи (рисунок 8). Поскольку открытие челюстей вызывает повышение первой форманты, то сопрано при пении увеличивает ее открытие синхронно с частотой фонации, обеспечивая при этом сдвиг – подстройку первой формантной области к частоте фонации. На рисунке 9 показано, как сдвигаются первые три форманты в женском оперном пении при увеличении высоты тона (величины фундаментальной частоты). Эта методика используется для большинства гласных в верхнем регистре сопрано, а также в верхнем диапазоне тенорами. Открытие губ не увеличивает эффективность излучения, так как источник не производит больше энергии, но это сдвигает форманты, что может изменить перераспределение энергии в определенных диапазонах и улучшить помехозащищенность.

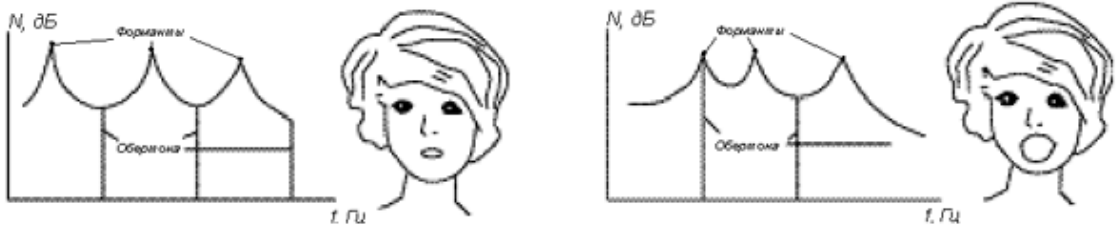
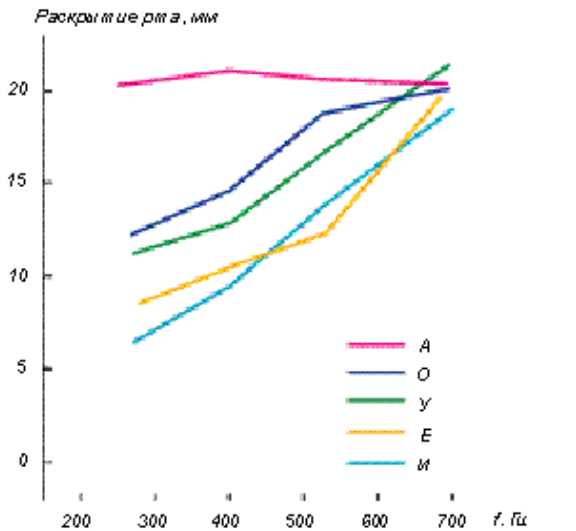


Рис. 9 Подстройка формант

Подстройка формант под первую фундаментальную частоту позволяет получить выигрыш в звуковом давлении, максимальное значение которого составляет 30 дБ. Это усиление звука является, как и певческая форманта, чисто резонансным процессом. Оно происходит без увеличения затрат энергии от вокального источника (колебаний голосовых связок). Выигрыш в громкости необходим также в связи с пением в

сопровождении оркестра, других певцов или аккомпанемента.

Для большинства гласных значение первой форманты лежит в области 300...800 Гц. Отсюда получается, что гласные, у которых первые форманты ниже 500 Гц, могут маскироваться оркестром. Если фундаментальная частота превышает 500 Гц (В4), и происходит подстройка формант (сближение, кластеризация формантных частот), то общая громкость усиливается, поскольку первая и вторая форманты, где имеет место максимальное усиление звука, сдвигаются по частотному диапазону и сближаются друг с другом, и высокие женские голоса хорошо слышны даже на фоне громкого оркестра. То есть достигается такой же эффект, как и в мужских и низких женских голосах, где для этих целей используется другой механизм— певческая форманта.

Модификация качества гласных

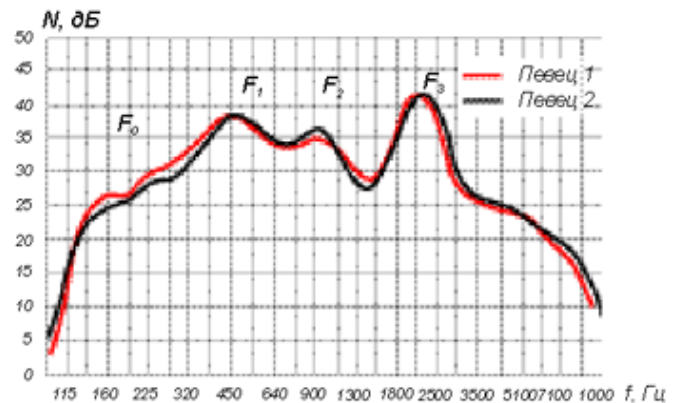
В пении для повышения формантной частоты при повышении высоты тона используется еще один прием, получивший название "модификация гласных".

Пение как с подстройкой формант, так и с певческой формантой, требует перестройки вокального тракта: расширения глотки, снижения гортани и др., что, естественно, влияет на две низшие формантные частоты, которые критичны для идентификации гласных. Сравнение всех формантных частот при речи и пении (рисунок 3) показывает, что вторая и третья форманты значительно различаются (вторая форманта при пении ниже по частоте), что оказывает влияние на разборчивость гласных. Поэтому при пении используется прием модификации гласных (например, "э" заменяется на "а" и т.д.) для того, чтобы перейти на гласные звуки, формантные области которых ближе к тем, которые формируются при вокальной речи с высокой певческой формантой.

Интегральные характеристики вокальной речи

Рис. 10 Три области максимумов на огибающей вокальной речи

Если произвести статистическую обработку спектров вокальной речи большого количества певцов с их индивидуальными формантами, то в среднем получаются кривые, показанные на рисунке 10. На них отчетливо видны три области максимумов (чего нет на обобщенной спектральной кривой для речи). Наибольших уровней достигает в мужских голосах третья область — это высокая певческая форманта, о которой уже было сказано выше.

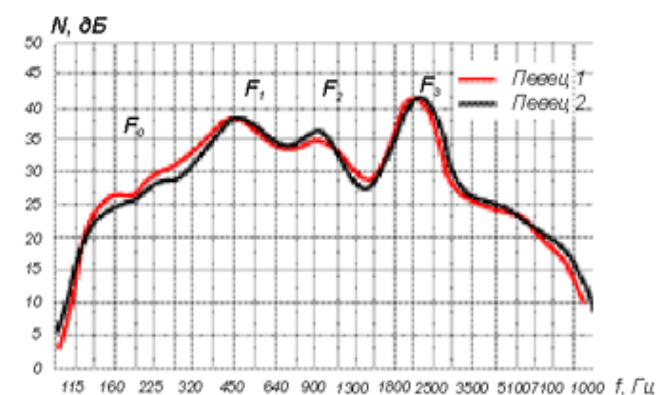


Первую область называют первой (нижней) певческой формантой, она определяет мягкость, массивность голоса. Вторая область, которую называют артикуляционной (фонетической) формантой, ниже по уровню и совпадает со второй формантой звука "а" (хотя певец поет все гласные, по-видимому, при пении они все становятся ближе к "а" по артикуляции). Различия между басами, баритонами и тенорами отчетливо видны на расположении этих трех максимумов:

	F1, Гц	F2, Гц	F3, Гц
Басы	380...540	760...1100	2100...2500
Баритоны	450...540	1100	2500
Тенора	540...640	1300	2500...3000

Рис. 11 Огибающая для женских голосов

В женских певческих голосах также отчетливо видны три стабильные области максимумов (рисунок 11), только основная частота на октаву выше, чем у мужчин, и практически совпадает с первой формантной областью (примерно 450 Гц), вторая форманта выражена сильнее, чем у мужчин, а третья область 3,5 кГц ниже по уровню.



В обычной речи при произнесении различных гласных формантные области все время смещаются в процессе речевого потока, эта вариабельность и обеспечивает разборчивость речи. Одним из требований к хорошему певческому голосу является "ровность" тембра, так как, если при пении произносить гласные как и при речи, то имеет место "пестрота" пения. Это требование приводит к выравниванию гласных — при пении на высоких тонах все гласные воспринимаются как "а", так как из-за подстройки формант все гласные поются с широко открытым ртом. По-видимому, в начальный период атаки звука все форманты устанавливаются для опознания звуков, затем голосообразующий аппарат перестраивается для пения и стабилизирует форманты. Это, конечно, не сглаживает общие тембральные различия при исполнении разных произведений разными людьми, поскольку сохраняются существенные различия в высших обертонах, характере атаки и спада звука и др. Однако эту специфику спектральных

характеристик вокальной речи следует учитывать при ее обработке и записи.

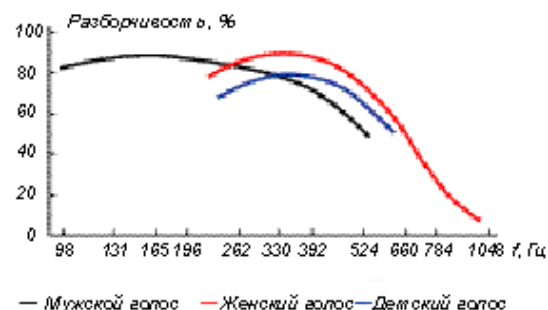
Разборчивость вокальной речи

Спектральная перестройка голоса для повышения его помехоустойчивости как с помощью создания особой певческой форманты, так и с помощью подстройки формант, неизбежно приводит к перестройке первых формантных областей по частоте, амплитуде, добротности и др., а поскольку именно первые две формантные области отвечают за распознавание гласных в речи, то перестройка первых формантных областей в процессе пения приводит в целом к ухудшению разборчивости вокальной речи. Хотя главная задача вокальной речи состоит в передаче эмоционально-эстетической информации, но передача семантической информации (понимание произносимого текста) также имеет существенное значение (во всяком случае, в классической вокальной культуре). Многие педагоги, певцы (например, Карузо, Шаляпин и др.) и композиторы уделяли очень большое значение дикции при исполнении вокальных произведений (хотя дикция – более широкое понятие, включающее в себя как разборчивость, так и другие особенности произношения).

Попытки количественно оценить разборчивость вокальной речи и влияние на нее перестройки формант были предприняты в работах В.П. Морозова. За основу был взят субъективный метод оценки слоговой разборчивости ("Звукорежиссер", 5/2002). Группа тренированных экспертов также проводила прослушивание слогов из стандартных артикуляционных таблиц, произносимых как в обычной речи, так и пропетых на разной высоте тона. Статистическая обработка большого количества испытаний позволила выявить некоторые общие закономерности: средняя разборчивость вокальной речи оказалась у всех испытуемых вокалистов ниже, чем разборчивость обычной речи у них же – ее значение снизилось в среднем с 87 до 70,7%. Причем, если у мужчин эти величины составляли 86,7 и 75%, то у женщин – 88% и 65%, то есть разборчивость женских голосов в вокальной речи падает гораздо значительнее, что, очевидно, связано с высотой их голоса.

Рис. 12 Зависимость разборчивости вокальной речи от высоты тона

Зависимость слоговой разборчивости от высоты основного тона голоса показана для мужских, женских и детских голосов на рисунке 12. Как следует из полученных данных, область максимальной разборчивости соответствует частоте основного тона для мужских голосов в области 110...262 Гц (почти полторы октавы), а для женских голосов – 330...524 Гц. При повышении частоты основного тона до 500 Гц разборчивость мужских голосов падает, но все-таки сохраняется на уровне примерно 50%, в женских голосах при повышении основного тона до 104 Гц (С3) разборчивость практически стремится к нулю.



Таким образом, зона хорошей разборчивости находится в средней части диапазона голоса, несколько уменьшается в области низких частот и значительно падает в области высоких, причем у женских голосов значительно больше, чем у мужских.

Поэтому в области средних частот, совпадающей с речевым диапазоном для данного певца, разборчивость вокальной речи практически такая же, как и обычной речи, но по мере повышения высоты голоса разборчивость резко падает. Существует несколько основных причин этого явления. При повышении частоты основного тона (например, до 1000 Гц) на высоких тонах остается слишком мало обертонов, которые попадают в слышимый диапазон, это также меняет тембр звуков и затрудняет их опознание. Основной тон оказывается в области за пределами формантных частотных областей, и певец старается произвести подстройку формант, но сдвиг первых формант резко ухудшает распознавание звуков речи и соответственно разборчивость. Более 30% энергии концентрируется в области певческой форманты, соответственно снижается уровень первых формант, что также сказывается на разборчивости. Наконец, в вокальной речи особенно проявляется разница по уровню и по длительности между гласными и согласными звуками. Уровень гласных может достигать 110...120 дБ, уровень согласных 40...70 дБ, что приводит к явлениям маскировки согласных, а, поскольку согласные передают основную семантическую информацию, то значительная потеря разборчивости в вокальной речи обусловлена также нечеткой артикуляцией согласных, что важно учитывать при записи и обработке речи.

В целом, если ставится задача передачи слова в пении, то лучше, если произведение написано для средних (и низких) регистров, что всегда учитывалось раньше композиторами (по-видимому, на интуитивном уровне); при пении в высоких регистрах передача вербального содержания происходит плохо, основная задача здесь – передача эмоций.

Часть 17. Слух и речь, часть 5 Акустические характеристики вокальной речи (продолжение)

В предыдущей статье были рассмотрены резонансные акустические характеристики вокальной речи, к числу которых относятся: образование особой певческой форманты, способы подстройки формант и др., что обуславливает особые интегральные акустические характеристики вокальной речи и создает проблемы с ее разборчивостью. В этой статье будет продолжен анализ акустических параметров, в первую очередь, таких важных для исполнительского и звукорежиссерского творчества, как образование вибрато и тремоло, а также использование различных регистров и связанных с этим изменений основных характеристик звукового сигнала.

Одним из средств музыкальной выразительности в пении является использование в процессе исполнения различных видов *модуляции* (изменения) параметров звукового сигнала. Такие изменения воспринимаются как тембровая особенность исполнения, придающая ему живость, полетность и т. п. Модуляция параметров широко используется не только в пении, но и при исполнении музыки на различных инструментах: струнных, духовых и др.

Как известно из физики, в простом гармоническом колебании возможно использование трех видов модуляции: амплитудной, частотной, фазовой. Модулированные колебания получаются в том случае, если какие-либо параметры сигнала (амплитуда, частота, фаза) начинают изменяться по заданному закону. Процесс управления параметрами звукового сигнала и называется его *модуляцией*. Простое гармоническое колебание полностью описывается тремя параметрами - амплитудой A_0 , частотой ω_0 и фазой φ_0 :

$$X(t) = A_0 \sin(\omega_0 t + \varphi_0).$$

Изменяя по определенному закону каждый из этих параметров, можно получить различные виды модуляции сигналов. Если меняется амплитуда сигнала, то этот процесс называется амплитудной модуляцией, если частота - частотной модуляцией, фаза - фазовой модуляцией.

Амплитудная модуляция (АМ): если происходит изменение амплитуды, то коэффициент A_0 будет уже не постоянным, а начнет меняться во времени $A(t)$. Это изменение можно производить, например, по гармоническому (синусоидальному) закону: $A(t) = A_0 \sin(\omega_m t + \varphi)$, где ω_m - частота изменения амплитуды, φ - фаза ее изменения. Такой вид зависимости амплитуды сигнала от времени называется модулирующим колебанием; на практике его частота обычно много ниже частоты основного колебания (модулируемого):

($\omega_m \ll \omega_0$). В этом случае общее результирующее колебание может быть записано в следующем виде: $X(t) = A_0 (1 + M \sin(\omega_m t + \varphi)) \sin(\omega_0 t + \varphi_0)$, где M называется коэффициентом модуляции, он равен отношению $M = A_1/A_0$, где A_1 - амплитуда модулирующего сигнала, A_0 - амплитуда исходного модулируемого сигнала. Величина M характеризует глубину амплитудной модуляции, то есть показывает, насколько амплитуда основного колебания меняется со временем, что воспринимается на слух, как небольшое изменение громкости.

Вид амплитудно-модулированного колебания при разных коэффициентах модуляции показан на рисунке 1. В музыкальном исполнительстве используются только сигналы с очень малым коэффициентом модуляции и низкой частотой изменения амплитуды, порядка 6...8 Гц. Такой вид изменения сигнала называется *тремоло*. В спектре модулированного таким образом сигнала имеется основной сигнал с частотой ω_0 и появляются два дополнительных сигнала с частотами $(\omega_0 + \omega_m)$ и $(\omega_0 - \omega_m)$ и амплитудами $A_0 M/2$.

Например, если несущий сигнал 1000 Гц промодулирован сигналом с частотой $\omega_m = 10$ Гц, то появятся две боковые полосы 1010 Гц и 990 Гц. Таким образом, спектр АМ-сигнала расширяется по полосе, и ширина его равна $2 \omega_m$ Гц. Если модулирующий сигнал содержит много гармоник, то спектр модулированного сигнала имеет вид, представленный на рисунке 2.

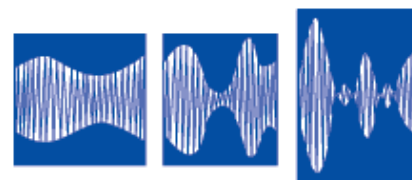


Рис. 1 Вид АМ-сигнала при разной глубине модуляции

Частотная модуляция (ЧМ): если в простом гармоническом колебании изменять по определенному закону частоту или фазу колебаний, то можно получить частотно- и фазомодулированные сигналы (ЧМ и ФМ). Если изменение частоты выполнить также по простому синусоидальному закону, то суммарное частотно-модулированное колебание будет иметь вид, показанный на рисунке 3, и может быть записано в виде (если считать $\varphi_0 = 0$): $X(t) = A_0 \sin(\omega_0 t + m \sin(\omega_m t))$, где $m = \Delta\omega/\omega_0$, называется индексом однотоновой частотной модуляции, $\Delta\omega$ - девиация (изменение) частоты исходного сигнала. В спектре такого сигнала также появляются боковые составляющие, не только пропорциональные частоте модуляции, но и ее гармоникам $(\omega_0 + n\omega_m)$, $(\omega_0 - n\omega_m)$. Структура спектра становится сложнее, с ростом индекса модуляции m увеличивается число спектральных составляющих, при этом происходит перераспределение энергии между ними, и расширяется ширина спектра, которая примерно равна $2(m+1)\omega_m$ Гц, то есть ЧМ-сигнал требует полосы частот больше в $(m+1)$ раз, чем АМ-сигнал, что имеет принципиальное значение (например, в радиовещании). Частотно-модулированные сигналы широко используются в вокальном и музыкальном творчестве и называются *вибрато* (обычно с частотой модуляции 5...10 Гц). На слух такая модуляция воспринимается как небольшие изменения высоты

тона.

При пении излучается сложный музыкальный многокомпонентный сигнал, и процесс модуляции его параметров представляет сложный взаимосвязанный процесс. Спектральный анализ записей оперных певцов показывает, что чаще всего в их пении используется частотная модуляция, пример показан на рисунке 4, то есть **вибрато**. Вибрато сопровождается и небольшой амплитудной модуляцией. В некоторых других видах пения (поп-музыке, народном пении и др.) иногда используется амплитудная модуляция, то есть **тремоло** (хотя в некоторых работах этот вид модуляции называется интенсивностное или амплитудное вибрато, а под термином "тремоло" понимается особый вид вибрато с большой частотой модуляции, мы будем придерживаться далее первого определения).

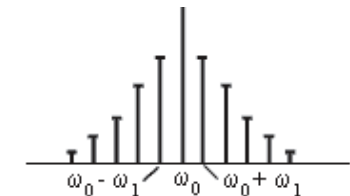


Рис. 2 Спектр АМ-сигнала

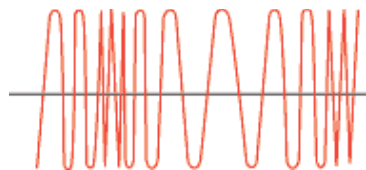


Рис. 3 Вид ЧМ-сигнала

Анализ особенностей модуляции в голосах выдающихся певцов, выполненный еще в 30-е годы в работах Сишора, затем в 80-е годы в работах Морозова, Сандберга, в современных работах Прейма и др. позволили выявить некоторые общие закономерности:

- в пении выдающихся оперных мастеров используется в основном частотная модуляция с частотой 5...7 Гц и глубиной модуляции ± 50 центов (что соответствует изменению от 31 до 98 Гц, в зависимости от расположения на частотной шкале).
- важной особенностью вибрато мастеров оперного вокального искусства (Т. Руффо, М. Баттистини, Ф. Таманьо, А. Патти, Н. Обухова, В. Собинов, И. Козловский, С. Лемешев и др.) является равномерная *периодичность* вибрато. Эта особенность иллюстрируется на рисунке 4. Форма модулирующей волны близка к синусоидальной, и даже если она отклоняется от синусоидальной, все равно сохраняется ее четкая ритмичность. А вот для голосов непрофессиональных певцов характерен неритмичный характер вибрато, что воспринимается как неустойчивость, неуверенность пения.
- значение частоты модуляции является характерной особенностью данного певца, хотя она (также как и глубина модуляции) может слегка видоизменяться в зависимости от высоты основного тона, силы голоса, типа гласной, длительности ее звучания и ряда других факторов. Существует также некоторая зависимость используемой частоты модуляции от типа голоса (рисунок 5).
- особо следует отметить, что параметры вибрато могут изменяться в зависимости от эмоциональной настроенности певца: при повышенном эмоциональном тоне при исполнении отдельных вокальных произведений частота вибрато в некоторых пределах возрастает.

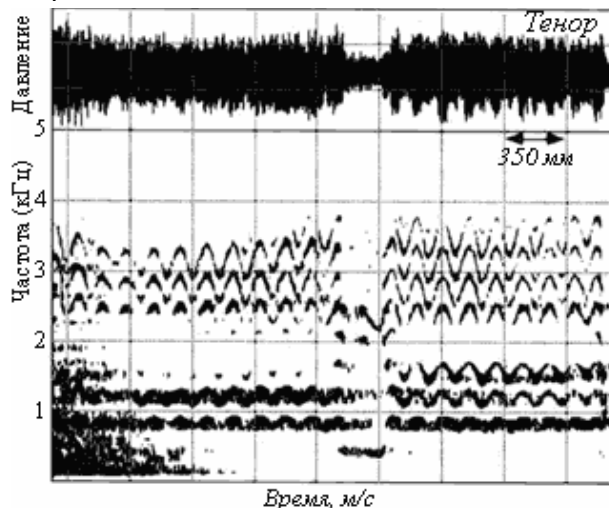


Рис. 4 Вибрато при оперном пении

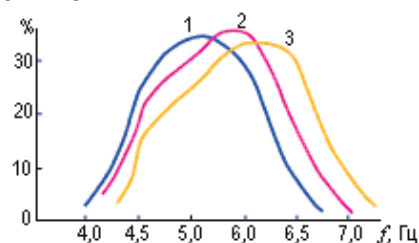


Рис. 5 Зависимость частоты модуляции от типа голоса

Использование частотной модуляции характерно в основном для оперного и камерного исполнения. В современной рок-музыке, где исполнение идет с огромным напряжением связок, модуляция практически не возникает, звук получается "прямым".

Механизм создания модуляции в вокальной речи (пении) исследовался с помощью электромиографии, т. е. измерения электрической активности мышц, которые показали, что в такт с колебаниями фундаментальной частоты, которые происходят с частотой модуляции, происходят сокращения мышц гортани: щитовидно-перстневидных (см. 1/2002), которые отвечают за удлинения связок и, соответственно, за изменение высоты тона. Кроме того, происходит изменение активности вокальных мышц и боковых черпаловидно-перстневидных мышц, отвечающих за разведение связок. Именно сокращения этих мышц служат причиной появления частотной модуляции. При этом, поскольку в колебаниях участвуют и мышцы сведения связок, то в такт с частотой модуляции меняется и сила сведения связок, а это приводит к изменению подглоточного давления, которое служит причиной изменения амплитуды сигнала, что и порождает соответствующую амплитудную модуляцию.

Существует и другая причина создающая сопровождающую амплитудную модуляцию при вибрато: при колебаниях голосовых связок происходит модуляция потока воздуха, и получаемый при этом сигнал имеет гармонический спектр, достаточно богатый обертонами (рисунок 6). При вибрато происходит изменение значений фундаментальной частоты, при этом, синхронно с фундаментальной, варьируются и значения частот всех обертонов. Амплитуда обертона зависит от того, как далеко он находится от формантной области,

при этом частоты формант при вибрато не смещаются. Поэтому при смещении частоты обертонов с частотой модуляции происходит увеличение или уменьшение их амплитуды в зависимости от того, дальше или ближе они перемещаются к формантной частоте. Поскольку общая громкость излучаемого сигнала зависит, в основном, от амплитуд обертонов внутри формантных областей, то происходит изменение громкости сигнала с частотой модуляции. Таким образом, при частотной модуляции, то есть использовании вибрато при пении, происходит и амплитудная модуляция сигнала, хотя она не так значительна.

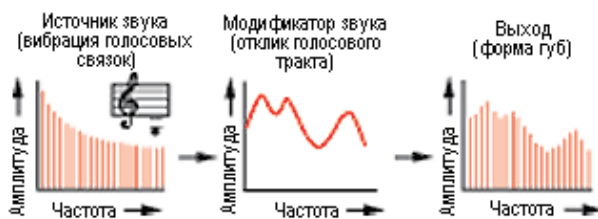


Рис. 6 Спектральный состав звука при колебаниях голосовых связок

Следует отметить, что существует специальный вид амплитудной модуляции, вызванный варьированием подглоточного давления, который воспринимается на слух как изменение громкости (тремоло) и применяется иногда в некоторых видах народного пения, но не используется в оперном или концертном исполнении.

Основные преимущества использования вибрато заключаются, прежде всего, в улучшении эстетического восприятия пения - звук кажется богаче (за счет обогащения спектра при модуляции), живее и т. д. Кроме того, существуют и объективные причины использования этого приема при пении, основные из которых следующие:

- применение вибрато уменьшает требования к точности настройки фундаментальной частоты, т. е. высоты тона. Исследования показали, что хотя при модуляции меняется основная частота, а, следовательно, и высота тона, но если частота модуляции лежит в пределах 6...7 Гц при глубине модуляции ± 40 центов, то слух как бы усредняет значения частоты, и воспринимает некоторую среднюю высоту тона. Если частота модуляции ниже 5 Гц, становятся слышны колебания высоты ("качания"), голос кажется очень напряженным (старческим). Если частота модуляции выше 7 Гц, голос воспринимается нервным, "дрожащим". Таким образом, диапазон изменения частоты модуляции 6...7 Гц оказывается оптимальным;
- использование вибрато уменьшает требования к точности настройки интервалов, и помогает скрыть биения, возникающие при неточной настройке голосов при полифоническом исполнении. К биениям слух очень чувствителен, порог чувствительности к расстройке частоты составляет меньше трех центов;
- применение вибрато повышает помехозащищенность голоса, поскольку максимальная чувствительность слуха к частотно-модулированным сигналам совпадает с областью 5...7 Гц, а это позволяет лучше выделять этот сигнал на фоне помех;
- вибрато повышает разборчивость звуков, особенно при пении в высокой тесситуре, поскольку смещение обертонов по частоте приводит к увеличению или уменьшению их амплитуд в зависимости от их позиции по отношению к формантным частотам. Это помогает также идентифицировать формантные области, а именно они служат основой распознавания гласных звуков.

Таким образом, выбор частоты модуляции в диапазоне 5...7 Гц оказался наилучшим образом согласованным с возможностями слуховой системы. Максимальная чувствительность слуха к этой области частот модуляции развилась, по-видимому, в процессе восприятия речи, которая представляет собой частотно- и амплитудно-модулированный процесс. С другой стороны, именно эти частоты модуляции легче всего могут быть реализованы мышечным аппаратом человека, в частности, общий резонанс дыхательной системы человека равен 6 Гц.

Использование модулированных (меняющихся во времени) звуков свойственно практически всем видам исполнения, поскольку они учитывают особенность организации слуховой системы человека, в которой при монотонном звучании выключаются нейроны, реагирующие на изменение звуковых сигналов ("нейроны новизны"), и у слушателя теряется интерес к исполнению. А модуляция сигналов поддерживают его внимание.

Кроме того, как считает Сандберг (один из самых известных в мире специалистов по акустике речи и пения, руководитель центра при Шведской Академии музыки), пение с вибрато сигнализирует слушателю, что процесс происходит легко, без фонационных проблем, а это усиливает его эстетическое восприятие.

Следует отметить, что поскольку вибрато (и тремоло) являются средствами для выделения голоса (привлечения к нему внимания), то использование его в хоровом пении может создать проблемы.

Поэтому певцы, воспитанные в сольной манере пения, не всегда вписываются в ансамбль.

В целом, поскольку наличие модуляции в голосе является дополнительным средством художественной выразительности, а современные возможности компьютерных технологий позволяют создать практически любые виды амплитудной, частотной и фазовой модуляций, то этим средством могут достаточно широко пользоваться звукорежиссеры в процессе обработки вокальной речи. Правда, всегда надо ориентироваться на естественные параметры модуляции, иначе голос будет звучать неестественно.

Специфической особенностью вокальной речи является использование различных певческих **регистров** - в обычной речи они используются только в случаях крайних эмоциональных состояний. В идеале голос должен

производить все доступные ему звуки без резких изменений в тембре, но на практике это недостижимо. По поводу определения и классификации регистров в научной литературе имеются противоречивые мнения. Наиболее распространенным является следующее определение: "Регистр - фонационный ряд частот, в котором все звуки звучат одинаково по тембру (при этом, по-видимому, они и создаются подобным образом)". Переход от одного регистра к другому отражает изменения в фонационной технике, то есть изменения взаимодействия мускул гортани и голосовых связок. Смена регистров ощущается как сдвиг в фонационной частоте (высоте тона), и как изменение тембра. При этом происходит и изменение других акустических характеристик звукового сигнала: изменение формы спектральной огибающей, структуры атаки и спада и др. Поскольку смена регистров используется в пении как средство художественной выразительности, то анализ происходящих при этом изменений акустических характеристик может оказаться полезным в практике работы звукорежиссера при обработке голоса.

Для классификации и названий регистров в разных вокальных школах используются совершенно разные понятия, в итальянской школе, например, выделяется три регистра: грудной (chest), средний (middle), головной (head) в женском голосе, и грудной (modal, normal), головной и фальцет (falsetto) в мужском голосе, со звуковысотными диапазонами, показанными на рисунке 7. Обычно принято считать, что механизмы голосообразования сильно различаются в нижних ("тяжелых") и верхних ("легких") регистрах; средние регистры сочетают качества их обоих. Басы и контральто поют почти исключительно в тяжелых регистрах, используя легкие регистры иногда на высших нотах (например, для комического эффекта). Колоратурные сопрано поют в легких регистрах, иногда используя "тяжелый" механизм на низших нотах - но никогда не поют только в грудном регистре. Иногда дополнительно выделяются в самом верхнем диапазоне женского голоса "свистящие" регистры, или очень низкие регистры в мужском голосе (так называемый "strobass").



Рис. 6 Спектральный состав звука при колебаниях голосовых связок

Отдельные регистры перекрываются по частоте, и некоторые певцы могут петь одни и те же ноты в разных регистрах. Например, в женских голосах области перекрытия: грудной-средний (chest-middle) 400 Гц, нота "соль" первой октавы, средний-головной (middle-head) 660 Гц, нота "ми" второй октавы. В мужских голосах зона перекрытия 200...350 Гц.

Главная задача вокальной педагогики - сделать переход от одного регистра к другому плавным, уменьшить вариации тембра так, чтобы этот переход был почти незаметен при восприятии. У неопытных певцов иногда происходит перелом - срыв в фальцет.

Механизм образования регистров лежит в изменении свойств голосового источника - гортани и голосовых связок. На колебания связок оказывают влияние аэродинамические факторы (величина подглоточного давления), мышечные (миоэластические) факторы, зависящие от натяжения основных видов мышц (вокальных, щитовидно-перстеновидных, щитовидно-черпаловидных и др.), а также упруго-механические свойства всех трех слоев самих голосовых связок.

В простейшем приближении частоты колебаний голосовых связок определяются так же, как для обычной струны:

$$f=0,5\sqrt{\frac{T}{l\cdot m}}$$

Как следует из этой формулы, для увеличения частоты фонации необходимо увеличивать натяжение (**T**), однако при этом меняется длина голосовых связок (**l**) и уменьшается их поверхностная плотность. В целом, при увеличении натяжения частота колебаний голосовых связок возрастает.

Таким образом, частота фонации зависит от натяжения и массы (**m**), поэтому при пении на низких частотах связки расслабленные, толстые и короткие, а на высоких - тонкие, длинные и напряженные. Повышение основной частоты почти линейно связано с удлинением связок.

Можно выделить два вида натяжения в голосовых связках - *внутреннее* и *внешнее*, при изменении которых происходит изменение массы, упругости и общей формы связок, то есть их длины, толщины, ширины и т. д.

Внутреннее натяжение определяется сжатием вокальных мышц, находящихся непосредственно внутри голосовых связок - при этом связки становятся короче и жестче, и за счет этого меняется фундаментальная частота их колебаний.

Внешнее натяжение происходит за счет межперстневидно-щитовидных мышц, которые при вращении щитовидного хряща относительно перстневидного (рисунок 8) обеспечивают как натяжение, так и удлинение голосовых связок. Взаимодействие этих двух видов мышц, а также черпаловидных мышц обеспечивает как изменение фундаментальной частоты, так и изменение тембра звучания, то есть переход к разным регистрам.

В "тяжелых" и "легких" регистрах механизм взаимодействия мышц гортани и соответственно формы колебаний голосовых связок существенно различаются.

В грудном регистре вокальные мышцы активны, они делают связки толстыми и короткими, натяжение связок при этом сравнительно небольшое:

- связки колеблются целиком всей массой и плотно прилегают друг к другу, закрытие и открытие вокальных связок начинается у нижней поверхности, пока сила давления не разомкнет их по всей толщине, при этом возникает вдоль толщины волна смещения;
- длительность их смыкания превосходит время размыкания. Вид модулированного потока воздуха показан на рисунке 9;
- периодические резкие толчки воздуха дают полный звук, богатый обертонами, амплитуда которых слабо убывает с частотой, придавая голосу богатый оттенок. Спектр звука в этом регистре показан на рисунке 6;
- энергия, подводимая от легких, довольно эффективно преобразуется в звуковую энергию, то есть КПД преобразования энергии достаточно высокое;
- моды колебаний связок в этом регистре показаны на рисунке 10. Как видно из рисунков, в связках возникают разные формы (моды) колебаний в трех разных плоскостях: полуволновой резонанс в горизонтальной плоскости, полуволновой в вертикальной плоскости, и четвертьволновой вдоль толщины связок.
- при таком звуке на передней грудной клетке происходят сильные вибрации, поэтому его называют "грудным".

Таким образом, в тяжелых регистрах превалируют мышечные факторы внутри голосовых связок, при этом потребляемый объем воздуха должен быть достаточно большим.

Исследования показали, что пока вокальные мышцы активны, голос держится в модальном грудном регистре. Когда начинает ослабевать напряжение вокальных мышц, голос начинает переходить в легкие регистры.

Легкий регистр (фальцет, головной) характеризуется тем, что в нем голосовые связки колеблются иначе: внутренние вокальные мышцы расслаблены, натяжение определяется в основном внешними мышцами, связки становятся длиннее примерно на 30%, и тоньше, меньше масса на единицу длины:

- продольное напряжение в связках сравнительно высокое, причем во всех трех слоях, амплитуда их вибраций довольно мала, в колебаниях принимает участие только внешняя кромка связок, голосовая щель только сужается-расширяется, происходит ослабление-усиление струи воздуха, а не полное ее прерывание, как при грудном регистре;
- закрытие голосовой щели короткое и неполное из-за большого натяжения связок, больше 70% цикла связки остаются открытыми. Это приводит к изменению модуляции воздушного потока, и более "дыхательному" звучанию голоса. (рисунок 10) ;
- фальцетный регистр не богат обертонами, спектр показан на рисунке 11. Колебаний грудной клетки почти нет;

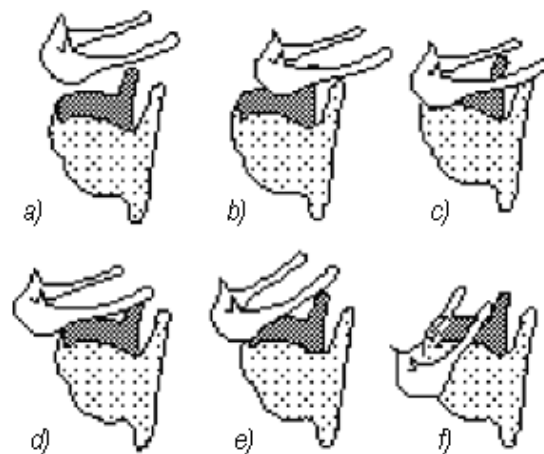


Рис. 8 Работа гортани при смене регистров

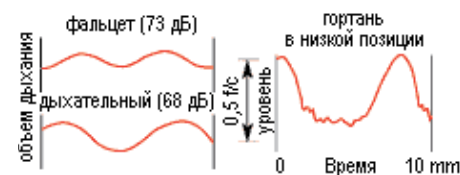


Рис. 9 Модуляция потока воздуха в грудном регистре и фальцете

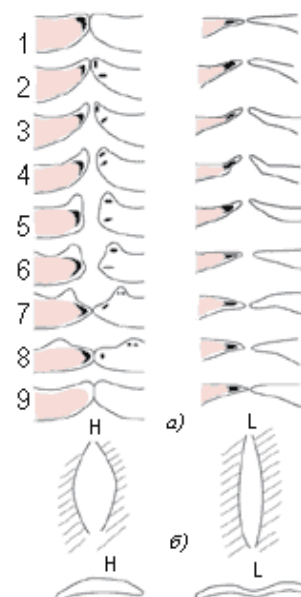


Рис. 10 Формы колебаний голосовых связок в грудном регистре и фальцете

- конверсия легочной энергии в звуковую энергию менее эффективна, и звуки не такие громкие;
- формы колебаний связок в этом регистре показаны на рисунке 10: в горизонтальной плоскости полуволновой резонанс, резонанс по толщине отсутствует, в вертикальной плоскости 3/4-волновой резонанс (второй).

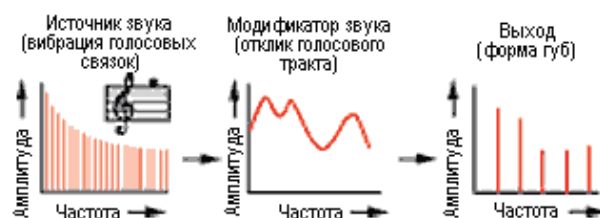


Рис. 11 Спектр звуков в высоком регистре

При дальнейшем повышении высоты звука голосовая щель укорачивается и подвигается к передней половине голосовых связок, ближе к черпаловидным хрящам, при этом задняя половина связок остается плотно сжатой и в колебаниях не участвует - связки имеют вид Y. Такие частичные колебания связок возможны благодаря особому строению голосового мускула.

В этом диапазоне частот начинают превалировать аэродинамические факторы, мускулы уже не могут сокращаться при дальнейшем повышении высоты тона. Воздух, проходя мимо сильно напряженных связок, создает турбулентный шум, поэтому иногда эту часть диапазона называют свистящим регистром. Образование звуков в этом случае происходит аналогично тому, как это имеет место в лабиальных духовых инструментах (флейте) за счет краевых тонов. Певцы редко пользуются только грудным или только фальцетным регистром, и избегают быстрых переходов от одного к другому. Одним из средств незаметного перехода от грудного регистра к фальцетному служит механизм смешанного голоса (средний регистр), когда характерное для грудного регистра напряжение вокальных мышц начинает ослабевать, и заменяется постепенно увеличивающимся напряжением внешних растягивателей. Ширина этого регистра разная для разных типов голосов. При пении резкая смена тембров воспринимается как изменение регистра. Таким образом, смена регистров - сложный процесс, связанный в первую очередь с изменением свойств голосового источника: формой колебаний голосовых связок, изменением взаимодействия различных мышечных структур гортани, формой модуляции воздушного потока и др.), который требует отработки при постановке голоса.

Однако, учитывая специфику изменений акустических характеристик, происходящих при смене регистров, звукорежиссер в процессе обработки вокала может исправить нежелательные изменения звукового сигнала, приводящие к изменению тембра.