

# АКУСТИЧЕСКИЙ АСПЕКТ ФОНЕТИЧЕСКИХ ОПИСАНИЙ

## Предмет акустической фонетики

§ 49. В акустической фонетике изучаются аэродинамическая и акустическая фазы речи:

- *аэродинамическая* фаза – сжатие и выталкивание воздуха, возникновение воздушных импульсов, завихрений вследствие движений артикулирующих органов;
- *акустическая* фаза – колебания воздушного давления (звуковые волны), которые являются следствием аэродинамических процессов и распространяются от говорящего к слушающему.

Основной объект изучения в акустической фонетике – акустический речевой сигнал, который является единственной наблюдаемой формой речевого произведения. Акустический сигнал представляет собой результат работы механизма речеобразования, поэтому одна из важнейших задач акустической фонетики – изучение связи между артикуляцией и ее аэродинамическими следствиями. Кроме того, изучение акустического речевого сигнала позволяет узнать и о перцептивных свойствах звуков, так как именно он поступает на вход механизма речевосприятия. Сегодняшняя речевая акустика тесно связана и с чисто прикладными задачами – например, такими, как синтез и анализ речи при помощи компьютера.

## Физическая природа звука

§ 50. Объектом изучения в акустической фонетике является **речевой сигнал** – передающиеся по среде добавочные (по сравнению с атмосферным давлением) колебания<sup>1</sup> воздушного давления, которые возникают в результате аэродинамических и акустических процессов, обусловленных артикуляцией<sup>2</sup>. Это давление называется звуковым или акустическим. Изменения звукового давления представляют собой волны, которые могут распространяться в какой-либо среде – в газе, жидкости или твердом теле (их распространение невозможно лишь в вакууме); в дальнейшем речь будет идти только о распространении звуков в воздухе. В воздухе изменения плотности и давления, вызванные в одном месте, распространяются по всем направлениям от этого места; волны сжатия и разрежения, распространяющиеся в воздухе при колебаниях тел или внезапных изменениях его плотности, называются звуковыми волнами. Их можно рассматривать с временной точки зрения (в одной и той же точке пространства давление изменяется с течением времени) или с пространственной (в одно и то же время величина давления может быть разной в разных точках пространства).

Итак, с акустической точки зрения **звук речи** – это результат воздействия на слуховой аппарат человека колебательных движений воздушной среды, вызванных артикуляцией.

## Виды колебаний. Периодические и непериодические колебания

§ 51. Для создания любых колебаний необходим источник или сила, вызывающая колебания. Источник колебаний может быть различным: движения струны, поток воздуха через узкое отверстие, удар и т.п. Колебания по разным основаниям подразделяются на

- периодические и непериодические,
- простые и сложные (комплексные),
- свободные и вынужденные.

Важнейшее различие между колебаниями – наличие или отсутствие регулярно повторяющейся модели. В зависимости от этого выделяются колебания *периодические* (тон) и *непериодические* (шум). В случае периодических колебаний число колебательных циклов за время существования колебаний достаточно велико, а колебательные циклы не отличаются друг от друга. Примером периодических колебаний могут служить колебания струны, маятника, а в речи – голосовых связок<sup>3</sup>. Примером непериодических колебаний являются, например, колебания воздушного шарика на ветру (у них нет регулярного повторения модели) или колебания, вызванные ударом молотка по стеклу (у них нет повторения вообще, то есть число колебательных циклов слишком мало). Периодические колебания могут быть как простыми, так и сложными<sup>4</sup>, непериодические колебания всегда являются сложными.

<sup>1</sup> Колебания – это изменения состояния какого-либо тела, обладающие той или иной степенью повторяемости во времени. Звуковые колебания возникают при внезапном изменении плотности воздуха – появлении сгущений или разрежений в каком-либо месте воздушной среды.

<sup>2</sup> Оно приблизительно в миллион раз меньше атмосферного.

<sup>3</sup> Строго говоря, периодическими можно считать только такие колебания, которые, во-первых, длятся бесконечно, а во-вторых, имеют абсолютно идентичные колебательные циклы. Таких колебаний в природе не существует, так как, во-первых, все колебательные движения со временем затухают, а во-вторых, природные звуки имеют не вполне идентичные колебательные циклы. Поэтому, говоря о периодических колебаниях, мы имеем в виду колебания, которые в действительности являются **квазипериодическими** – то есть такими, которые хотя и затухают, но достаточно медленно (по сравнению со временем одного колебательного цикла), а структура их периодов в значительной мере сходна, хотя и не идентична.

<sup>4</sup> О простых и сложных, а также о свободных и вынужденных колебаниях см. ниже.

## Объективные свойства звуков и их субъективные корреляты

§ 52. При изучении речевых звуков необходимо описывать их объективные характеристики и субъективные свойства, то есть соответствующие ощущения, возникающие при их восприятии человеком. К числу объективных характеристик звуков относятся частота, интенсивность, спектр и длительность, к числу субъективных свойств – соответственно высота, громкость, тембр и долгота.

**Частота (f)** – это число колебательных движений, полных колебательных циклов в единицу времени. Одно колебательное движение в одну секунду равно одному герцу, сокращенно Гц. Молодой человек в состоянии различать частоты от 16 до 18000-20000 Гц., а, например, летучие мыши используют диапазон от 20000 до 100000 Гц. Особенно важны для речи частоты от 100 до 5000-8000 Гц., именно в этом диапазоне сосредоточен максимум информации о речевых звуках – так, например, мы практически без потерь можем воспринимать речь даже по телефону, который пропускает сигнал в диапазоне от 300 Гц. до 3500 Гц.

Субъективное восприятие частоты называется *высотой*. Единицей измерения высоты является *мел*.

Время, в течение которого совершается один полный колебательный цикл, называется *периодом* колебания (T). Период является величиной, обратной частоте колебания:  $f = 1/T$ . Если частота колебания составляет 20 Гц., то период равен 1/20 секунды.

**Амплитудой** колебаний называется величина максимального изменения звукового давления.

**Интенсивность (i)** или сила звука определяется его мощностью. Мощность звука – это энергия, которая излучается источником в единицу времени (измеряется в ваттах, Вт), а интенсивность или сила звука – это мощность звуковой волны, которая приходится на площадку  $1 \text{ м}^2$  (перпендикулярную направлению распространения волны). Интенсивность измеряется в  $\text{Вт}/\text{м}^2$ , но в этих абсолютных физических величинах ее представляют редко, так как диапазон звуковых интенсивностей, доступных восприятию человека, чрезвычайно велик: интенсивность самых сильных звуков, воспринимаемых человеком (порог болевого ощущения), в  $10^{14}$  раз больше интенсивности самых слабых (порога слышимости). Эта разница очень велика, поэтому для измерения интенсивности используется логарифмическая шкала децибел. В соответствии с этой шкалой, изменению силы звука в 10 раз соответствует изменение на 10 дБ., в 100 раз – на 20 дБ., в 1000 раз – на 30 дБ. и т.д. На расстоянии 1 метра от источника звука значение интенсивности в 20 дБ. приблизительно соответствует шуршанию листьев, 30 – шепоту, 45 – шуму пишущей машинки, 60 – обычному разговору, 75 – пению или крику, 100 – шуму движущегося поезда метро, 120 – шуму взлетающего самолета.

**Громкость** звука (субъективное восприятие интенсивности) зависит не только от интенсивности, но и (в значительно меньшей степени) от частоты колебаний: более высокие звуки при той же интенсивности воспринимаются как более громкие.

**Спектр** звука – это относительная амплитуда всех его частотных составляющих (подробнее о спектре см. ниже §§ 55-57). Субъективное восприятие спектра называется *тембром*.

**Длительность** речевых звуков измеряется в миллисекундах (мс.) и составляет обычно не менее 25 мс. В среднем же длительность звуков речи находится в пределах 40-200 мс. Субъективно звуки могут восприниматься как долгие и краткие.

## Распространение звуковых волн

§ 53. Скорость звука (с) в воздухе составляет около 350 м/сек. или 1260 км/ч. Скорость звука относительно постоянна<sup>5</sup> и не зависит от его интенсивности – громкие и тихие звуки "путешествуют" с одинаковой скоростью (но громкие дальше, так как интенсивность звука обратно пропорциональна квадрату расстояния от источника). Ступение или разрежение воздуха, возникшее около источника звука, с течением времени распространяется в пространстве. Если источник звука – колеблющееся тело, звуковая волна за время, равное периоду колебаний тела T, успевает пройти расстояние, равное произведению скорости звука на длительность периода. Это расстояние называется **длиной звуковой волны** (см. рис. 10) и обозначается греческой буквой "лямбда" ( $\lambda = c * T$ ). Поскольку  $T = 1/f$  (см. выше § 52), то эту формулу можно записать в виде  $\lambda = c/f$ , то есть длина волны прямо пропорциональна скорости распространения волн в данной среде (с) и обратно пропорциональна частоте колебаний (f).

## Простой (чистый) тон – гармоническое колебание

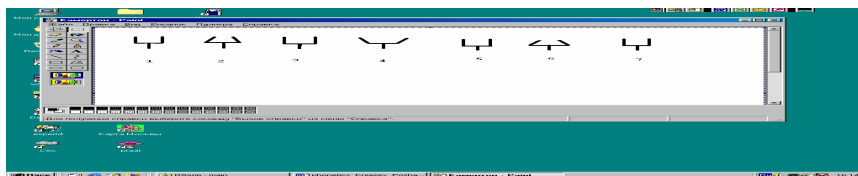
§ 54. Речевые звуки представляют собой комплексные колебания, т.е. сложнейшие сочетания простых или чистых тонов и/или шумов.

**Простой тон** – это периодическое колебание, которое имеет только одну частоту колебания. Иначе простое периодическое колебание называется *гармоническим*.

Звуков такого рода в природе не существует, хотя имеются звуки очень близкие чистому тону. К ним относится, например, звук, издаваемый камертоном. Если ударить по стеблю камертона, то его ножки начинают смещаться из нейтрального положения, затем возвращаются в исходное положение под воздействием силы эластичности, затем, вследствие инерции, продолжают движение через точку покоя, затем обратно и т.д. (см. рис. 1.2, 1.3; 3.2, 3.8). Силы инерции и эластичности противоположны и действуют в любой момент движения, при этом то одна сильнее, то другая.

<sup>5</sup> Она зависит от температуры и влажности воздуха.

Рисунок 11. Схематическое изображение смещения усов камертона за полтора колебательных цикла. Положение 1 – состояние покоя; положение 2 – смещение внутрь под действием внешней силы, действие силы эластичности; положение 3 – возвращение в состояние покоя, действие силы эластичности уменьшается, а силы инерции увеличивается; положение 4 – смещение наружу, действие силы эластичности увеличивается, а силы инерции уменьшается; положение 5 – возвращение в состояние покоя, действие силы эластичности уменьшается, а силы инерции увеличивается (конец первого колебательного цикла); положение 6 – смещение внутрь, действие силы эластичности увеличивается, а силы инерции уменьшается; положение 7 – возвращение в состояние покоя, действие силы эластичности уменьшается, а силы инерции увеличивается.



Движение камертона вызывает движение окружающих его молекул воздуха, которое можно сравнить с колебанием обыкновенных качелей (см. рис. 13). Движущиеся молекулы вызывают движение соседних молекул (как бы "подталкивают" их – см. рис. 14), в результате образуются последовательные сгущения и разрежения воздуха – звуковые волны. Звуковые волны распространяются концентрическими кругами, как волны от камня, брошенного в воду: сжатия и разрежения воздушной среды чередуются (см. рис. 15). Эти чередования давления во времени (в одной и той же точке) могут быть представлены в виде графика (осциллограммы)<sup>6</sup>, на котором время откладывается по горизонтальной оси, а давление – по вертикальной (см. рис. 16). Графиком простого периодического (гармонического) колебания является синусоида.

Вследствие действия силы трения точки наибольшего смещения частиц воздуха все больше приближаются к точке покоя: амплитуда колебания уменьшается, происходит затухание колебания (*damping* – см. рис. 17 и Б10), однако частота колебаний (количество полных циклов в единицу времени) остается постоянной.

Гармонические колебания могут различаться по частоте, амплитуде и фазе

Одна и та же среда может передавать множество звуков одновременно. При этом колебания (например, при наличии нескольких источников) могут взаимодействовать друг с другом. Если их частота совпадает, то амплитуда просто суммируется (и это по-прежнему простой тон)<sup>7</sup> (см. рис. 18а).

## Комплексные звуки. Спектральное разложение Фурье

§ 55. Большинство источников производят не простые, а сложные (комплексные) колебания, то есть колебания, характеризующиеся наличием более чем одной частоты. Так, например, при колебании струны колеблется не только вся она целиком, но и ее половина, четвертая часть и т.п. Все речевые звуки являются сложными.

Комплексные колебания могут быть 1) (квази)периодическими, т.е. имеющими регулярно повторяющуюся модель и 2) непериодическими (у них отсутствует регулярное повторение модели вследствие неповторяющегося характера колебаний или быстрого их затухания).

В периодических звуках частоты всех составляющих их колебаний кратны самой низкой частоте – частоте колебания всего тела, которая называется частотой основного тона (сокращенно – ЧОТ, обозначается  $F_0$ ). **Частота основного тона** – это частота повторения полных колебательных циклов в единицу времени<sup>8</sup>. Если вся струна колеблется с частотой 100 Гц., то в то же самое время ее части колеблются с частотами 200, 300, 400 и т.д. Гц. Все частоты, кратные ЧОТ, называются *гармониками* (включая саму ЧОТ) или *обертóнами* (все гармоники, кроме ЧОТ)<sup>9</sup>.

График комплексного колебания – не синусоида, а гораздо более сложная линия: в этом случае молекулы среды колеблются не в одном направлении, как маятник, а по гораздо более сложной траектории (см. рис. 19).

<sup>6</sup> Осциллограмма получается следующим образом: при помощи микрофона звуковые колебания преобразуются в электрические, а при помощи осциллографа они регистрируются и представляются в графической форме.

<sup>7</sup> Впрочем, результат зависит еще и от фазы колебаний (синфазные тоны одной частоты и амплитуды воспринимаются как единый звук с удвоенной громкостью, а находящиеся в противофазе (сдвиг по фазе  $180^\circ$ ) подавляют друг друга – см. рис. 18б, 18с).

<sup>8</sup> Частота основного тона голоса зависит от 1) напряженности связок, которая регулируется сокращением мышечной части связок (увеличение напряженности повышает ЧОТ) или подъемом перстневидного хряща (поэтому передние гласные более высокие); 2) перепада давлений по обе стороны глоттиса: повышение подвязочного давления приводит к повышению ЧОТ (более громкие звуки являются и более высокими).

<sup>9</sup> Количество гармоник в спектре звука не ограничено, но их амплитуда уменьшается с увеличением частоты (на 12 дБ. каждую октаву).

Чтобы получить комплексный периодический звук, можно просто сложить несколько простых тонов (см. рис. 19 и рис. Б11 в Приложении Б). Верно и обратное – сложный тон можно разложить на несколько простых. Такое разложение называется спектральным анализом.

Спектральный анализ сложных звуков сводится к получению значений частот и амплитуд простых тонов (гармоник), из которых состоит комплексный звук. Это преобразование<sup>10</sup> основано на **теореме Ж.Б.Фурье**: *любое периодическое колебание является суммой гармонических колебаний (простых тонов), частоты которых кратны частоте основного тона, причем эти гармонические колебания различаются по фазе и амплитуде.*

Визуальное изображение амплитуд гармоник называется *амплитудно-частотным спектром*, при этом амплитуда гармоник откладывается по вертикальной оси, а частота – по горизонтальной (см. рис. 20). Значения амплитуды на спектре часто не калибруются, т.к. важны не абсолютные ее значения, а относительная амплитуда гармоник. Линия, соединяющая вершины гармоник амплитудно-частотного спектра, называется *спектральной огибающей*.

В непериодических звуках нет регулярного повторения модели, однако амплитудно-частотный спектр можно получить и для них .

## Резонанс

§ 56. Итак, одной из причин сложности большинства звуков является наличие в их источнике нескольких частотных составляющих. Другая причина связана с явлением резонанса, в результате которого изменяются амплитуды гармоник источника. Именно резонансы, возникающие в речевом тракте, позволяют получить огромное разнообразие звуков речи при наличии весьма ограниченного набора источников речевых звуков: так, например, все гласные образуются в результате действия одного (голосового) источника звука.

В самом упрощенном виде, резонанс – это вибрация одного тела, вызванная вибрацией другого.

Все колебания подразделяются на *свободные и вынужденные*. У всех тел есть собственная частота колебаний<sup>11</sup>. Для того, чтобы какое-либо тело начало колебаться, нужно вывести его из состояния покоя – воздействовать на него тем или иным способом – например, подтолкнуть качели или подбросить мяч. Если это воздействие осуществляется однократно, то дальнейшие колебания тела называются свободными. Частота этих колебаний совпадает с собственной частотой колеблющегося тела. Свободные колебания достаточно быстро затухают (их амплитуда постоянно уменьшается и постепенно доходит до нуля: качели, если их больше не раскачивать, быстро останавливаются, а мяч, если его не трогать, перестает подпрыгивать).

Чтобы колебания не затухали, их можно поддерживать какой-либо внешней силой, которая будет добавлять энергию и предотвращать уменьшение амплитуды. Колебания, поддерживаемые внешней силой, называются вынужденными. В случае вынужденных колебаний собственная частота колеблющегося тела заменяется частотой колебаний вынуждающего тела – качели будут раскачиваться не с собственной частотой, а с той частотой, с которой их подталкивают. Если же эти частоты (собственная частота колеблющегося тела и частота вынуждающей силы) совпадают и вынуждающая сила действует "в такт" с колебаниями данного тела, амплитуда колебаний резко возрастает: так, даже если мы будем подталкивать качели совсем слабо, но в такт их собственным колебаниям, качели можно раскачать очень сильно. В этом и состоит явление резонанса:

**Резонанс** – это увеличение амплитуды вынужденных колебаний, вызванное совпадением частоты вынуждающей силы и частоты колеблющегося тела.

**Резонатором** называется тело, колебания которого вызваны другими колебаниями. Резонатор сам не создает колебаний, а только усиливает те колебания, вызванные источником, частота которых совпадает с его собственной частотой (или приближается к ней). Так, струна усиливает те колебания камертона, частота которых совпадает с ее собственной.

Струна, качели – это примеры механических резонаторов. Для речевой акустики гораздо большее значение имеют *акустические резонаторы* – контейнеры с воздухом, определенный объем которого тоже является резонатором. Примером действия акустического резонатора может служить резонанс, возникающий при наливании воды в бутылку: в этом случае звук становится все выше и выше, так как объем воздуха уменьшается (для акустических резонаторов, одним из которых является речевой тракт человека, наибольшее значение имеет объем, меньшее – форма).

Для того чтобы скрипка, виолончель или гитара звучала, недостаточно только струн определенной длины, толщины и натяжения – нужен резонатор, усиливающий определенные частоты (более высокие у скрипки, чем, например, у виолончели). Точно так же и при речеобразовании в резонаторах речевого тракта – полостях глотки, рта и носа – усиливаются некоторые частоты источника звука (в зависимости от объема и формы этих полостей). В этом состоит основное положение акустической теории речеобразования: **свойства речевых звуков определяются свойствами их источников и резонаторов речевого тракта.**

<sup>10</sup> Оно называется спектральным разложением Фурье.

<sup>11</sup> Любое тело может вибрировать и резонировать, но способность к вибрации у разных тел, конечно, разная (например, струна, один раз выведенная из состояния покоя, может вибрировать сама по себе довольно долго, а мембрана телефонной трубки – только в том случае, если ее постоянно вынуждать, подталкивать).

## Акустическая теория речеобразования

§ 57. Итак, с точки зрения акустики речевой тракт представляет собой совокупность соединенных между собой воздушных резервуаров (полостей глотки, рта и носа), каждый из которых (и все они вместе) может служить резонатором. Чтобы заставить колебаться столб воздуха, заключенный в этих резервуарах, необходимо воздействовать на них какой-либо силой. Эта сила создается *источником* звука, который может находиться как в самом речевом тракте, так и за его пределами (таким источником являются, например, голосовые связки)<sup>12</sup>.

Существует два основных типа источников речевых звуков:

1. *голосовой* (квазипериодический) – для звонких звуков, и
2. *шумовой* (непериодический) – для согласных. Шумовой источник может быть
  - импульсным (в случае смычных согласных) – в этом случае колебания создаются резким скачком давления в результате раскрытия смычки и быстро затухают;
  - турбулентным (с его участием образуются фрикативные согласные) – в этом случае колебания создаются воздушной струей, возникшей вследствие наличия сужения в речевом тракте, и могут поддерживаться длительное время.

При произношении некоторых звуков имеется только один источник: голосовой у гласных, импульсный – у глухих взрывных, турбулентный – у глухих фрикативных. Возможна, однако, и любая комбинация источников (т.е. при произношении одного звука их может быть сразу несколько): например, голосовой и турбулентный источники участвуют в образовании звонких фрикативных согласных; голосовой и импульсный – звонких смычных; турбулентный и импульсный – глухих аффрикат; голосовой, турбулентный и импульсный источники необходимы для образования звонких аффрикат.

В акустической теории речеобразования (АТР), создателем которой является шведский ученый Гуннар Фант, **речь** рассматривается как процесс **фильтрации**. Это означает, что речевой тракт выступает в функции фильтра, пропускающего (усиливающего) только те частоты, порожденные источником звука, которые совпадают с его собственной частотой (см. рис. 21, 22). Надгортанные полости являются резонаторами, собственные частоты которых могут довольно значительно изменяться в зависимости от положения артикулирующих органов, придающих им разный объем и форму (наибольшее значение имеет при этом длина резонатора, а также место и площадь его поперечного сечения).

Если обозначить через *S* спектр источника (англ. source – "источник"), через *T* – спектральную характеристику фильтра, в роли которого выступает речевой тракт (англ. transfer – "передача"), и через *P* – спектр результирующего звука, то акустическую характеристику звука речи можно представить равенством:  $P = S \cdot T$ . Это равенство является формальной записью **основного положения акустической теории речеобразования**: *спектр звука является результатом воздействия одного или нескольких источников на фильтрующую систему речевого тракта*.

## Форманта. F-картина

§ 58. Итак, в акустической картине (спектре) звука наиболее усиленными оказываются те частотные области, которые совпадают с частотами резонатора. Эти частотные области называются формантами. **Форманты** – это резонансные частоты речевого тракта определенной формы и объема. Частоты формант (кроме частоты основного тона) задаются, в первую очередь, конфигурацией речевого тракта<sup>13</sup>, что позволяет соотнести их с определенными целевыми артикуляциями и по частотам формант судить о положении артикулирующих органов.

Форманты обозначаются буквой *F*; их нумерация начинается с нулевой форманты – частоты основного тона ( $F_0$ ), далее следуют первая ( $F_1$ ), вторая ( $F_2$ ), третья ( $F_3$ ) и четвертая ( $F_4$ ) форманты. Совокупность значений формант называется **F-картиной** (формантной картиной).

При образовании некоторых звуков кроме резонансов в речевом тракте могут возникать и антирезонансные явления. Антирезонансы резко ослабляют амплитуду составляющих с частотами, близкими частоте антирезонанса, что приводит к подавлению близких резонансных частот или образованию глубоких (часто до нуля) минимумов в спектре – антиформант (нулей)<sup>14</sup>.

<sup>12</sup> Таким образом, речевой тракт всегда выступает в качестве резонатора и часто – как источник звука (в речевом тракте может образовываться шум).

<sup>13</sup> На практике же часто используется другое понимание этого термина: форманты – это гармоники, усиленные в результате резонанса, характеристика сигнала, результат модификации источника при помощи фильтра (а не свойство самого речевого тракта). Однако частотные значения формант далеко не всегда совпадают со значениями гармоник, особенно при большом значении ЧОТ, при котором гармоники расположены достаточно редко.

<sup>14</sup> Появление антирезонансов связано с сильной расчлененностью резонаторной системы или образованием в ней замкнутых воздушных полостей (такие конфигурации речевого тракта характерны для согласных и носовых гласных)

## Основные способы изучения акустических свойств речи

§ 59. Исследование акустических свойств речи опирается на преобразование звуковых колебаний в электрические (при помощи микрофона), а затем – в визуальное изображение (при помощи спектрографа, осциллографа или заменяющей их компьютерной программы). Основные виды визуального представления звуков перечислены в Таблице 6.

Таблица 6. Основные виды визуального представления звуков.

|   | ось X   | ось Y     |
|---|---|-----------|
| осциллограмма   | время   | амплитуда |
| мгновенная спектрограмма (спектральный срез)                    | частота   | амплитуда |
| динамическая спектрограмма<br>(узкополосная или широкополосная) | время<br>(амплитуда передается степенью зачернения) | частота   |

*Осциллограмма*<sup>15</sup> позволяет измерять, в первую очередь, длительность звуков, но не их частотные составляющие.

Самый распространенный вид акустического анализа речевого сигнала – спектральный анализ, позволяющий определить относительные амплитуды частотных составляющих звука. Основным принципом спектрографии – использование фильтров, выполняющих функцию резонаторов для тех электрических колебаний, в которые при помощи микрофона преобразованы звуковые колебания. Из всего набора фильтров на подаваемый сигнал откликаются только те, собственная частота которых близка к частоте исследуемого звука (при этом отклик тем сильнее, чем более интенсивна данная частота).

В спектрографе весь диапазон речевых частот (50 – 10000 Гц.) разбит фильтрами на определенное число шагов. В зависимости от их числа полоса одного фильтра может быть различной, поэтому спектрограммы делятся на *узкополосные* и *широкополосные* (см. рис. 23). В узкополосных спектрограммах ширина полосы составляет 30-50 Гц., и на них можно наблюдать гармоники звука и даже изменения ЧОТ, однако центр формантной области довольно трудно найти, особенно, если две форманты расположены близко друг к другу, поскольку он может не совпадать ни с одной гармоникой. В широкополосных спектрограммах ширина полосы составляет 300-500 Гц. (обычно это более двух гармоник), на них достоверно отражаются и непериодические сигналы. Поскольку широкополосные фильтры возбуждаются гораздо быстрее, чем узкополосные<sup>16</sup>, то на широкополосных спектрограммах можно увидеть быстрые изменения сигнала, отсутствующие на узкополосных.

Если измерить значения частот в одной точке акустического сигнала<sup>17</sup>, можно получить *спектральный срез* (или мгновенную спектрограмму)<sup>18</sup>; для анализа изменений сигнала во времени используются *динамические спектрограммы* (измерения производятся через определенные временные интервалы).

Интенсивность колебаний при спектральном анализе регистрируется путем последовательного измерения напряжения на всех фильтрах. В результате получается информация об относительной интенсивности всех частотных составляющих (т.е., о спектре).

В последнее время спектральный анализ осуществляется при помощи компьютера: звуковой сигнал сначала преобразуется в электрический, затем в цифровую форму (это называется "оцифровкой" – см. ниже), затем – в изображение.

## Образование гласных звуков

§ 60. Как происходит образование гласных с точки зрения акустики? Колеблющиеся частицы воздуха создают волны (сжатия и разрежения среды), которые распространяются со скоростью  $c = 350$  м/с. Частота их колебаний зависит от длины волны:  $f = c/\lambda$ .

Речевой тракт при произнесении некоторых звуков (например, самого простого гласного [ъ]) можно представить в виде трубки, открытой со стороны губ и закрытой с другой стороны голосовыми связками. Длина этой трубки (L) составляет около 17,5 см., а диаметр – около 1 см.<sup>19</sup> (см. рис. 24). Как будет воздух вибрировать в этой трубке в ответ на колебания источника (голосовых связок)? Можно ли попытаться предсказать, какие именно частоты источника будут усиливаться?

Вибрация резонатора похожа на вибрацию динамика или пружины в трубке (см. рис. 25). При этом частицы воздуха у закрытого конца трубки имеют гораздо худшие условия для колебаний, так как находятся в ограниченном пространстве, поэтому амплитуда их колебаний будет меньше (а давление больше), чем у открытого конца трубки, где может происходить обмен с окружающим пространством. Если сопоставить эти вариации давления со звуковой волной (представить их как часть звуковой волны), то можно увидеть, какие частоты будут резонировать – совпадать с кривой скорости частиц или звукового давления в данной трубке.

<sup>15</sup> Осциллограммы слова Иванова можно увидеть в Приложении Г.

<sup>16</sup> Резонаторы могут реагировать либо в ответ на узкий ранг частот (в этом случае колебания медленно достигают максимума и медленно затухают), либо в ответ на широкий диапазон (в этом случае нарастание и затухание происходит быстро). Узкополосные резонаторы (или фильтры) очень долго возбуждаются и затухают, но гораздо точнее быстрых широкополосных.

<sup>17</sup> Точка – это условное понятие. В таких случаях обычно производится усреднение приблизительно за 30-50 мс.

<sup>18</sup> Спектральные срезы русских гласных приведены в Приложении Б (рис. Б1 – Б9).

<sup>19</sup> Заметим, что диаметр практически не имеет значения в том случае, если он значительно меньше длины.

Напомним, что частота колебаний определяется длиной волны ( $f=c/\lambda$ ), так как скорость распространения звука – величина постоянная. На рисунке 26 кривая скорости частиц продолжена до полного цикла. Как видно из рисунка, длина этой волны ( $\lambda$ ) в 4 раза больше длины трубки ( $L$ ):  $\lambda=4L=4 \cdot 17,5 \text{ см.} = 70 \text{ см.}$ , следовательно, частота первого резонанса и первой форманты составит  $f=c/\lambda=35000 \text{ см/сек.} : 70 \text{ см.} = 500 \text{ Гц.}$  Но это только один из возможных резонансов и далеко не единственный. Кривые, представляющие другие частоты, тоже могут вызывать резонанс при условии, что амплитуда колебаний частиц будет минимальной (равной нулю) у закрытого конца и максимальной у открытого, пусть и с более сложными вариациями в середине.

Из рисунков 27-28 видно, что этому условию удовлетворяют волны с длиной волны  $4/3 L$  ( $f=c/\lambda=35000 \cdot 3 : 70 = 1500 \text{ Гц.}$ ),  $4/5 L$  ( $f=2500 \text{ Гц.}$ ),  $4/7 L$  ( $f=3500 \text{ Гц.}$ ) и т.д., но другие резонансы будут гораздо слабее, поскольку амплитуда колебаний сильно уменьшается с увеличением частоты (3500 Гц. – это 7 раз по 500 Гц., то есть 7 октав, следовательно, амплитуда меньше на  $7 \times 6 = 42 \text{ дБ.}$ ). Таким образом,  $f=c(2n-1)/4L$ .

Так же можно вычислить и резонансы для других гласных, например, [а]. Здесь уже речевой тракт можно условно представить в виде двух трубок, одна из которых (от места сужения в речевом тракте<sup>20</sup> до губ) шире другой (от гортани до места сужения в речевом тракте). Поскольку каждая из этих трубок в 2 раза короче одной, то резонансы будут в области 1000 Гц., 3000 Гц. и т.д. Но поскольку в этом случае вторая трубка не совсем закрыта с одного конца, то первый резонанс первой трубки будет чуть меньше (900 Гц.), а первый резонанс второй – чуть больше (1100 Гц.). Для лабиализованных гласных нужно будет еще учитывать увеличение длины резонатора за счет вытягивания губ, что будет понижать частоты резонатора.

## Соотношение артикуляционных и акустических характеристик гласных

§ 61. Итак, звук речи – это результат модификации спектральных составляющих источника вследствие резонансов, возникающих в речевом тракте. Как уже говорилось, резонансы речевого тракта различной длины и формы называются формантами. Наиболее существенными характеристиками гласных являются значения первой и второй форманты, менее важной – положение третьей форманты<sup>21</sup>.

Частоты формант (кроме частоты основного тона) задаются, в первую очередь, конфигурацией речевого тракта, что позволяет соотнести их с определенными целевыми артикуляциями.

Значение первой форманты гласного прямо пропорционально величине выходного (ротового) отверстия (чем больше отверстие, тем выше первая форманта) и обратно пропорционально объему полости глотки (чем больше полость глотки, тем ниже  $F_1$ ). С артикуляционной точки зрения оба эти параметра в значительной степени определяются подъемом языка при артикуляции гласного. Значение второй форманты обратно пропорционально длине ротового резонатора, которая, в свою очередь, соотносима с рядом гласного.

Сравним для примера гласные [и], [а] и [у]. У гласных верхнего подъема [у] и [и] величина выходного отверстия невелика (у [у] совсем мала вследствие лабиализации – см. рис. 29, 30), а ширина полости глотки довольно велика (особенно у [и] – за счет смещения языка вперед), поэтому значение  $F_1$  у этих гласных является минимальным. Наоборот, у гласного нижнего подъема [а] (см. рис. 31) величина выходного отверстия максимальна, а ширина полости глотки минимальна, поэтому значение  $F_1$  максимально велико.

Что касается длины ротового резонатора (расстояния от места наибольшего сужения до губ), то оно является минимальным у переднего гласного [и] (и значение его второй форманты самое большое) и максимальным у заднего гласного [у], при произнесении которого длина ротовой полости еще увеличивается за счет вытягивания губ (результатом чего является наименьшее значение  $F_2$ ); гласный среднего ряда [а] в этом отношении занимает промежуточное положение (см. Таблицу 7).

Таблица 7. Зависимость значений первой и второй формант гласных от размеров выходного отверстия и полостей рта и глотки.

|                                     | [и]            | [а]            | [у]                  |
|-------------------------------------|----------------|----------------|----------------------|
| <i>величина выходного отверстия</i> | <i>малая</i>   | <i>большая</i> | <i>очень малая</i>   |
| <i>объем полости глотки</i>         | <i>большой</i> | <i>малый</i>   | <i>средний</i>       |
| $F_1$                               | <i>низкая</i>  | <i>высокая</i> | <i>низкая</i>        |
| <i>длина ротовой полости</i>        | <i>малая</i>   | <i>средняя</i> | <i>очень большая</i> |
| $F_2$                               | <i>высокая</i> | <i>средняя</i> | <i>очень низкая</i>  |

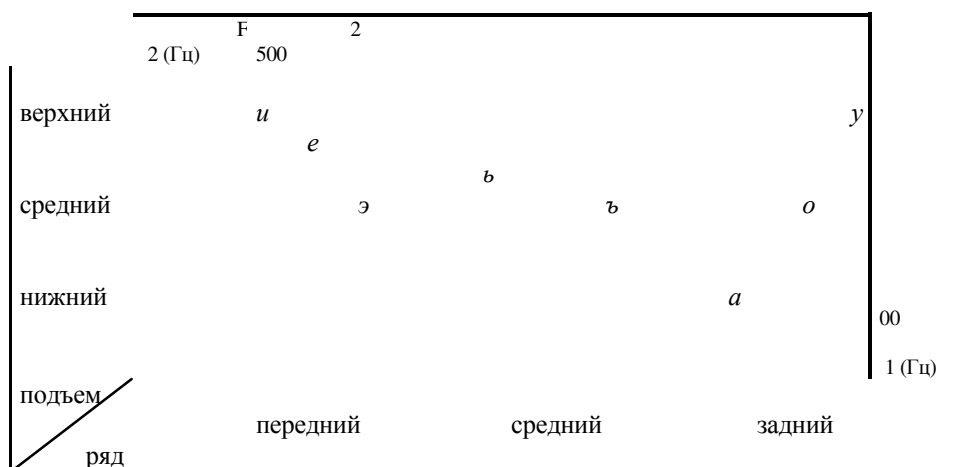
Абсолютные значения формант у разных людей могут варьировать в довольно широких пределах, но общая формантная структура (F-картина) с низкой  $F_1$  – высокой  $F_2$  для [и], низкими  $F_1$  и  $F_2$  для [у] и высокой  $F_1$  – средней  $F_2$  для [а] сохраняется всегда. F-картина гласного [е] очень похожа на F-картину [и] (хотя его первая форманта несколько выше, а вторая – несколько ниже), а F-картина [о] очень похожа на F-картину [у] (хотя обе его форманты несколько выше, чем у [у]). Гласный [э] по своей формантной структуре занимает среднее положение между [и] и [а].

<sup>20</sup> При артикуляции [а] сужение имеет место в полости глотки.

<sup>21</sup> Необходимо иметь в виду, что все гласные имеют и нулевую форманту – частоту основного тона, но она практически не зависит от свойств речевого тракта при произнесении конкретного гласного.

Соотношение артикуляционных (подъем и ряд) и акустических (значения  $F_1$  и  $F_2$ ) свойств гласных приведено на рисунке 32.

Рисунок 32. Соотношение значений первой и второй форманты гласных с их рядом и подъемом.



Гласный [ы], в отличие от всех других гласных, в произношении большинства носителей русского языка имеет дифтонгический характер. Это проявляется в том, что вторая форманта [ы] постепенно повышается, часто доходя до значения, характерного для [и] (2200-2500 Гц.). Гласный [ь], как уже упоминалось ранее, характеризуется значением  $F_1$  около 500 Гц. и  $F_2$  около 1500 Гц., гласный [ь] – соответственно 400 и 1700 Гц.

Ударные гласные отличаются от соответствующих безударных большей длительностью, а заударные от всех остальных – значительно меньшей интенсивностью.

Типичные спектральные срезы гласных, на которых можно увидеть их формантные картины, приведены в Приложении Б на рисунках Б1 – Б9. Форманты отражаются на этих рисунках в виде пиков.

Динамические спектрограммы гласных приведены в Приложении В на рисунках В1 – В6. На них форманты отражаются в виде горизонтальных черных полос.

§ 62. Увеличение общей длины резонатора и уменьшение выходного отверстия приводит к понижению всех формант, а ее уменьшение – к повышению. Сужение в передней части ротовой полости ведет к сближению  $F_2$  и  $F_3$  за счет повышения  $F_2$ , а сужение в задней части – к сближению  $F_2$  и  $F_1$  за счет понижения  $F_2$ . При уменьшении расстояния по частоте между двумя формантами их амплитуды увеличиваются, что приводит иногда к появлению общего максимума (это очень типично для  $F_1$  и  $F_2$  гласного [у]). Можно считать, что для трех основных гласных наиболее существенным является то, какие именно форманты сближаются: для [и] это  $F_2$  и  $F_3$ , для [а] и [у] –  $F_1$  и  $F_2$  (но в разных областях) – см. рис. 33.

В потоке речи один и тот же акустический эффект может быть достигнут при помощи различных артикуляционных стратегий. Так, при произнесении [у] губы часто не вытягиваются, а для достижения того же акустического эффекта язык отодвигается дальше назад (ротовая полость удлиняется), при этом – чтобы не уменьшался объем полости глотки – опускается гортань.

Традиционно классификация гласных строится на основании артикуляционных параметров – положения губ и языка во рту, хотя положение гласных в традиционном трапецоиде точнее соотносится со значениями  $F_1$  и  $F_2$ , чем с положением верхней точки языка на вертикальной и горизонтальной оси (ср. рисунки 9 и 32). Кроме того, следует отметить, что для артикуляции гласного положение высшей точки языка не так важно, как место наибольшего сужения во рту (так, например, для [а] наибольшее значение имеет сужение между задней частью языка и стенкой глотки).

§ 63. Гласные и сонорные согласные отличаются от шумных согласных наличием видимой формантной структуры (F-картины). У шумных согласных вследствие действия антирезонансов и возможного отсутствия голосового источника F-картина (или ее низкочастотная часть) может быть не видна (скрыта). В этом случае принято говорить о локусной формантной картине, которая может быть восстановлена по движению формант соседних сонорных звуков. **Локусы формант** – это те области в спектре шумных согласных, где должны находиться форманты, хотя они и не видны непосредственно на спектрограмме.

### Коартикуляционные изменения гласных

§ 64. В предыдущем разделе речь шла о формантной структуре изолированных гласных. Однако в реальной речи изолированное произнесение звуков практически не встречается, а соседние согласные (в большей степени предшествующие, чем последующие) очень сильно влияют на спектр гласных. В результате гласные практически всегда состоят не только из **стационарных участков** (где не наблюдается значительных изменений F-картины), но и из так называемых **переходных участков** (*transitions*), где эти изменения



происходят довольно быстро. Так,  $F_1$  гласного в соседстве со смычными согласными всегда понижается, а  $F_2$  – либо понижается, либо повышается в зависимости от ее значения на стационарном участке гласного и от значения локуса  $F_2$  согласного (он зависит от места образования согласного и его твердости/мягкости).

## Акустические свойства согласных

**§ 65. Сонорные согласные** по своей спектральной картине очень близки гласным и иногда отличаются от них только меньшей интенсивностью.

Боковой сонант [л] выделяется по очень высокой  $F_3$  и слабо выраженной низкой  $F_2$ . Дрожащий [р] лучше всего опознается по очень краткому (около 30 мс.) перерыву в звучании. Спектральная картина [ж] очень похожа на спектр [и], но может иметь и шумовые (непериодические) составляющие в верхней области частотного диапазона. Все сонанты (но особенно часто – мягкие) в значительной степени, а иногда и полностью, оглушаются в позиции конца слова или перед конечным шумным согласным.

У **носовых согласных**, в отличие от других сонорных, не два, а три резонатора (кроме полости рта и глотки это еще и полость носа). Открытый проход в носовую полость создает очень большой резонатор (глотка + нос) и, соответственно, сильный резонанс, обычно в области 200-300 Гц. (этот резонанс чуть выше у [н], у которого меньше ротовой резонатор). Кроме того, у носовых согласных наблюдается ослабление (антирезонанс) в области  $F_2$  соседнего гласного (у губных чаще всего в области 500-1500 Гц., у зубных – 2000-3000 Гц.).

### § 66. Шумные согласные.

Если у гласных резонансные полости находятся впереди источника звука (голосовых связок), то у шумных согласных они могут находиться как впереди, так и позади источника (шумообразующей преграды).

Резонансные полости, находящиеся впереди источника шума, оказывают гораздо более сильное влияние на спектр звука, чем резонансные полости, находящиеся позади источника.

Низкий локус  $F_1$  (значение которой связано с величиной ротового отверстия) является признаком смычного образования согласного, низкий локус  $F_2$  – губной артикуляции или дополнительной артикуляции веляризации; высокий локус  $F_2$  (в области около 2200-2500 Гц) свидетельствует о палатализации (или палатальности) согласного, ее средне-высокий локус (1500-1800 Гц) – о переднеязычной артикуляции.

**Взрывные** согласные выделяются по наличию паузы (отсутствию сигнала во время смычки), взрыву (резкому изменению спектральной картины) и послевзрывной фазе:

- у [п] она является самой непродолжительной, а максимум спектральной энергии не превышает 600 Гц.;
- у [т] максимум находится в высокочастотной (выше 1500 Гц.) области, но в целом энергия взрыва распределена практически по всему спектру;
- у [к] взрыв самый долгий (так как при его произношении больше всего расстояние от места артикуляции до губ) и самый интенсивный (так как давление воздуха за смычкой является наибольшим), чаще всего он расположен в области  $F_2$  следующего гласного (при его отсутствии – обычно ниже 2500 Гц.) и очень часто бывает двойным.

Впрочем, максимум информации о месте образования согласного можно получить не по его взрыву, а по переходному участку следующего гласного, содержащему информацию о локусной F-картине согласного: у велярных согласных наблюдается сближение  $F_2$  и  $F_3$ ; у твердых переднеязычных (кроме [л])  $F_2$  обычно расположена в области около 1500 Гц., а  $F_3$  бывает очень высокой; у губных все форманты имеют низкие значения. У всех звуков, соседних со смычными согласными, наблюдается понижение  $F_1$ .

**Щелевые** согласные характеризуются довольно долгим (100-200 мс.) шумовым периодом с плавным началом. При этом

- у [х] наблюдается самая низкая и узкая полоса усиленных частот (в той же области, что и у [к]),
- у [с] – самая высокая и широкая (там же, где у [т], наибольшее усиление энергии наблюдается в области выше 4000 Гц.),
- у [ш] – самый интенсивный шум (в области более низкой, чем у [с], максимум энергии ниже 4000 Гц.),
- у [ф] шум самый слабый, расположенный в нижней части спектра (там же, где у [п]).

Щелевые согласные, характеризующиеся интенсивным шумом, который образуется не в месте сужения, а в результате отражения воздушного потока разной скорости от преграды (зубов), называются *сибиллянтами* (зубные и переднеязычные).

**Аффрикаты** выделяются по наличию и смычки, и фрикативного шума (с резким началом, в отличие от постепенного у щелевых). Смычная и фрикативная части аффрикат короче, чем смычка и шум соответственно взрывных и щелевых согласных.

**Звонкие согласные** выделяются по наличию  $F_0$  и меньшей интенсивности взрыва и/или шума; кроме того, они несколько короче соответствующих глухих.

Сведения о **твердости/мягкости** согласных заключаются, в первую очередь, в переходных участках гласных: у всех гласных в положении рядом с мягким согласным  $F_2$  стремится в район 2200-2500 Гц.

Выше перечислены лишь самые основные сведения о спектральной картине различных звуков русского языка; при этом необходимо помнить, что эта картина в значительной степени зависит еще и от условий

произнесения и записи, а также от индивидуальных особенностей говорящего. Типичные динамические спектрограммы согласных приведены в Приложении В на рисунках В? – В?.

## Преобразование акустического сигнала в цифровую форму (оцифровка)

§ 67. В настоящее время анализ речи осуществляется, в основном, при помощи компьютеров, поэтому полезно понимать, как компьютер представляет звуковые волны. Он может делать это только в цифровой форме (в виде определенной последовательности цифр). Звуковую волну необходимо при этом преобразовать в серию чисел, соответствующих амплитуде колебаний в определенные моменты времени – через регулярные промежутки (см. рис. Б16 в Приложении Б). Частота измерения амплитуды сигнала называется частотой дискретизации (ЧД) – если амплитуда измеряется, например, каждые 0,2 мс. (5000 раз в секунду), то частота дискретизации равна 5000 Гц. Точность представления сигнала в цифровой форме зависит от количества измерений в единицу времени (сравни рис. Б16 и Б17 в Приложении Б). Если частота дискретизации недостаточна, то быстрые изменения амплитуды сигнала могут быть потеряны при оцифровке. Для ее приемлемого качества необходимо, чтобы, по крайней мере, все пики сигнала (как отрицательные, так и положительные) были отражены (ср. см. рис. Б18 в Приложении Б: сигнал, который содержит частоту 600 Гц., оцифрованный с ЧД 1400 и 1000 Гц. – 1-й хоть и не очень точно, но отражает все пики, 2-й – нет, т.к. они чаще, чем измерения амплитуды – и если попытаться воспроизвести сигнал по этим измерениям, то частота 600 Гц. будет потеряна; в 1-м же случае форма кривой несколько изменится, но все частотные составляющие сохранятся). При выборе ЧД для того или иного сигнала большое значение имеет теорема Кожевникова-Найквиста, гласящая: для адекватного отражения всех частотных составляющих сигнала частота дискретизации должна быть по меньшей мере в два раза выше его самой высокой составляющей. Частотой Найквиста называется, таким образом, частота, равная 1/2 ЧД. При этом очень важно, чтобы при оцифровке в сигнале не было частот выше частоты Найквиста, иначе не просто пропадают эти частоты, но и появляются новые, которых не было в исходном сигнале (артефакты, которые настолько же ниже частоты Найквиста, насколько были выше нее в исходном сигнале (см. рис. Б19 в Приложении Б); поэтому они должны быть отсечены до оцифровки при помощи фильтров.

Основная информация о речевых звуках сосредоточена в диапазоне от 0 до 8, максимум – до 11 КГц., значит, ЧД в 22 КГц. для их оцифровки вполне достаточно. Если же исследуются только гласные, то достаточно и 10 КГц, зато можно сэкономить много места в памяти компьютера. Оцифровка при записи компакт-дисков осуществляется с ЧД 44 000 Гц., цифровых кассет (DAT) – 48 000 Гц, что позволяет без искажений записывать все, что в состоянии различить человеческое ухо (а это не более 21 000 Гц.).

Частота дискретизации – не единственный фактор, от которого зависит качество сигнала, хранящегося в цифровой форме: важно еще, какова была амплитудная шкала при записи (разрядность преобразователя), то есть, сколько на ней было возможных точек (единиц). Чем подробнее эта шкала (чем больше на ней значений), тем точнее представление звука (см. рис. Б20, Б21 в Приложении Б). Компьютер может оперировать только целыми положительными числами и использует двоичную систему счисления, каждое число обозначается комбинацией нулей и единиц, а каждая цифра занимает 1 бит компьютерной памяти; так, для обозначения двойки или тройки нужно уже 2 бита, для цифр от 4 до 7 – 3 бита, от 8 до 16 – 4 и т.д. 8 бит составляют одно компьютерное слово (байт), так что 256 уровней амплитуды (0/255) требуют 8 бит (разрядов), а 65 536 – 16 бит (2 байта). Большинство компьютерных звуковых программ раньше использовали восьмиразрядные преобразователи, сейчас – 16 (CD/DAT) или даже 32-хразрядные (что означает наличие 131 072 уровней амплитуды). При использовании восьмиразрядного преобразователя соотношение между самым громким и самым тихим звуком в сигнале составляло всего 48 дБ., 16-тиразрядного – уже 96 дБ., 32-хразрядного – 192 дБ., т.е. больше того диапазона, который может различать человек (130 дБ.). Итак, чем больше частота дискретизации сигнала и разрядность преобразователя, тем лучше, однако при ЧД 44 000 Гц. и 16-тиразрядном преобразователе для записи одной секунды сигнала нужно 44 000 x 2 байта = 88 000 байт (5-7 мегабайт на одну минуту). Для того, чтобы просто записать такой текст в орфографии, нужно всего около 750 байт (в 7500 раз меньше).

## Алгоритм анализа спектрограмм

### § 68.

1. Найти пики sonorности и обозначить слоги.
2. Идентифицировать ударные слоги по относительной длительности и интенсивности.
3. Произвести сегментацию сигнала на акустические события – отметить резкие изменения амплитуды и/или частоты.
4. Приписать каждому сегменту все возможные признаки (слева направо или от простого к сложному): [р] (краткий перерыв) – сибиллянты (высокочастотный шум) – взрывные (пауза) и т.д. Использовать альтернативные решения: а) глухой-звонкий, б) шумный-сонорный, в) взрывной-фрикативный / гласный-полугласный-носовой, г) твердый/мягкий, д) губной-переднеязычный-велярный, е) ряд-подъем предударного и ударного гласных.
5. Выбрать дорожку сегментов, которая образует осмысленную цепочку.

## ЛИТЕРАТУРА.

- Бондарко Л.В.* Звуковой строй русского языка. М., 1977.  
*Динамические спектры речевых сигналов.* Под ред. Ф.М.Деркача. Львов, 1983.  
*Кодзасов С.В., Кривнова О.Ф.* Общая фонетика. М., 2001.
- Речь. Артикуляция и восприятие.* Под ред. Л.А.Чистович и В.А. Кожевникова. М.-Л., 1965.  
*Фант, Г.* Акустическая теория речеобразования. М., 1964.  
*Фланаган, Дж.* Анализ, синтез и восприятие речи. М, 1968.

# ПЕРЦЕПТИВНЫЙ АСПЕКТ ФОНЕТИЧЕСКИХ ОПИСАНИЙ

## Этапы процесса восприятия

§ 69. Восприятие является последним этапом коммуникации: слушающий воспринимает звуковой сигнал, переданный ему говорящим, и интерпретирует его определенным образом. Тем самым, в процессе восприятия происходит декодирование информации – преобразование ее из физической (акустической) формы в символическую (ментальную). Деятельность слушающего в процессе коммуникации направлена на понимание речевого сообщения.

В сложном процессе понимания речевого сообщения можно выделить несколько этапов<sup>22</sup>:

- прием акустического сигнала;
- преобразование сигнала (первичный слуховой анализ);
- выделение перцептивно значимых акустических событий и признаков;
- лингвистическая интерпретация звуковой стороны речевого сообщения (принятие решений).

1. На самом первом этапе происходит "прием" речевого сигнала барабанной перепонкой – "улавливание" и "регистрация" звуковых колебаний.

2. На следующем этапе осуществляется преобразование принятых колебаний в удобную для анализа форму, сходную с динамической спектрограммой речи – эта форма представления акустического сигнала называется "слуховой спектрограммой".

3. В центральных отделах слухового анализатора "слуховая спектрограмма" подвергается дальнейшей обработке, в процессе которой происходит выделение в ней наиболее существенных для восприятия акустических событий и признаков.

4. На завершающем этапе осуществляется лингвистическая интерпретация сигнала: выделенные ранее акустические события и признаки используются для принятия решений о фонемном составе принятого сообщения и о наличии в нем определенных знаковых единиц.

## Прием и преобразование акустического сигнала

### Устройство слухового аппарата

§ 70. Слуховой аппарат человека состоит из трех отделов:

- *периферического* (его образуют наружное, среднее и внутреннее ухо),
- *проводникового* (нервные пути) и
- *центрального* (находящегося в коре головного мозга).

### Периферический отдел слухового аппарата

§ 71. *Наружное ухо* состоит из ушной раковины и наружного слухового прохода, который представляет собой изогнутую трубку длиной около 2,5 см. и упирается в барабанную перепонку, отделяющую наружное ухо от среднего (см. рис. 34, 35). Поскольку слуховой проход – это полый сосуд, закрытый с одного конца и открытый с другого, то он является акустическим резонатором, в котором усиливаются определенные частоты (в диапазоне 2000-5000 Гц).

*Среднее ухо* отделено от наружного барабанной перепонкой. Барабанная перепонка – это вогнутая (площадью около 90 мм<sup>2</sup> и толщиной 0,1 мм.) кожно-волокнисто-слизистая структура, которая вибрирует целиком (в ответ на колебания низкой частоты) или частично (в ответ на колебания высокой частоты). В барабанную перепонку вращена рукоятка молоточка, который наряду с наковальней и стремечком<sup>23</sup> находится в среднем ухе или барабанной полости, наполненной воздухом (объемом около 1 см<sup>3</sup>).

От внутреннего уха среднее ухо также отделено перегородкой, в которой имеются два отверстия – круглое и овальное. Колебания барабанной перепонки приводят в движение молоточек, прикрепленный к ней одним своим концом; другой его конец соединен через наковальню со стремечком. Основание стремечка укреплено в овальном окне улитки – основного органа слухового отдела внутреннего уха. В системе среднего уха звуковое давление увеличивается приблизительно в 20 раз.

*Внутреннее ухо* (лабиринт) заполнено лимфатической жидкостью и имеет два отдела – слуховой и вестибулярный. Слуховой отдел внутреннего уха представляет собой полую костную спираль, свернутую приблизительно в 3 оборота, которая называется **улиткой**. Длина улитки составляет около 32 мм., а диаметр – 10 мм. у основания и менее 2 мм. у конца.

В улитке на границе со средним ухом есть два отверстия, закрытые эластичными мембранами. Одно из этих отверстий, овальное окно площадью около 3 мм<sup>2</sup>, соединено со стремечком, которое действует как

<sup>22</sup> см. С.В.Кодзасов, О.Ф.Кривнова. Общая фонетика. М., 2001.

<sup>23</sup> Молоточек, наковальня и стремечко называются слуховыми косточками.

поршень на жидкость, находящуюся внутри улитки. В результате жидкость приходит в движение и в ней создаются изменения давления<sup>24</sup>.

Внутри улитки по всей ее длине натянуты две перепонки (базиллярная мембрана и рейснерова мембрана), которые делят ее на три продольных канала (см. рис. 36, 37).

**Базиллярная мембрана** состоит из нескольких тысяч волокон, натянутых поперек улитки; эти волокна своими колебаниями реагируют на импульсы увеличения давления. Амплитуда колебаний разных участков мембраны различна и зависит от частоты возбуждающего тона и от амплитуды звукового сигнала. Разные места мембраны выделяют в сигнале колебания определенных частот<sup>25</sup>. Таким образом, базиллярная мембрана (аналогично спектрографу) как бы осуществляет преобразование Фурье, разлагая комплексные колебания на набор простых, различающихся по частоте и амплитуде. Например, звук [и] вызовет колебания разных участков мембраны, но самые сильные из них будут сосредоточены около ее конца (в ответ на колебания частотой 300 Гц.) и в середине (2500 Гц.).

В среднем канале улитки находится рецепторный аппарат слуха – *кóртиев орган*, который представляет собой сложную систему чувствительных *волосковых клеток-рецепторов*. Его функции состоят в том, чтобы обнаружить то место на базиллярной мембране, которое имеет максимальную амплитуду смещения, и передать соответствующий (электрический) импульс в головной мозг.

### **Проводниковый и центральный отделы слухового аппарата**

§ 72. Информация о колебаниях базиллярной мембраны, которые фиксируются с помощью волосковых клеток-датчиков, поступает в мозг через **слуховой нерв**. Этот нерв состоит из отдельных слуховых волокон, которые выходят из нервных клеток – первых слуховых нейронов. Каждый такой нейрон (а их у человека около 30000) имеет два волокна: одно подходит к рецепторным волосковым клеткам, а другое направлено в сторону мозга.

На выходе слухового нерва получается "изображение", которое показывает, как распределяется активность нейронов<sup>26</sup> в зависимости от их положения в слуховом нерве (то есть, в конечном счете, от частоты анализируемого звука). Это изображение называется **слуховым спектром** звукового сигнала.

По своей природе слуховой спектр аналогичен мгновенной спектрограмме. Изменение во времени распределения активности нейронов в частотных каналах создает внутреннее слуховое представление сигнала, подобное динамической спектрограмме.

Окончательная обработка полученного сигнала и принятие решений о звуковой структуре и содержании высказывания осуществляется в центральном отделе слуховой системы, который находится в коре головного мозга.

Как видно из изложенного выше, слуховая система человека, включающая разные структуры уха и центральный отдел, устроена и функционирует чрезвычайно сложным и тонким способом. В самом схематическом виде ее организация может быть представлена следующим образом:

### **Полезные признаки звукового сигнала (акустические ключи)**

§ 73. Слуховой аппарат человека устроен таким образом, что человек лучше всего слышит именно речевые сигналы (это и естественно, так как в процессе эволюции речевой и слуховой аппараты развивались параллельно).

Восприятие в целом опирается на более общую способность человека к поиску и распознаванию глубинных моделей. В случае восприятия речи эти модели – акустические, поэтому одной из основных задач перцептивной фонетики является поиск *акустических "ключей"* – тех акустических характеристик, которые человек использует для соотнесения того или иного отрезка сигнала с определенным звукотипом. Ключи эти часто бывают избыточными (то есть, о принадлежности сегмента к той или иной фонетической категории может свидетельствовать несколько разных фактов), что позволяет человеку осуществлять коммуникацию и в неблагоприятных условиях (например, при наличии помех, когда ряд полезных признаков оказывается недоступен).

Основной метод выделения акустических ключей заключается в проведении экспериментов по восприятию синтезированных сигналов, в которых варьируются определенные акустические параметры.

<sup>24</sup> Второе отверстие, круглое окно, служит для снижения избыточного давления.

<sup>25</sup> Различные участки базиллярной мембраны отличаются шириной и жесткостью. Ширина мембраны увеличивается по направлению от основания к вершине улитки примерно в десять раз, а упругость постепенно уменьшается. В соответствии со своим строением мембрана реагирует на разные частотные составляющие звуков колебаниями разных своих частей (см. рис. 38). Можно сказать, что место на базиллярной мембране является своеобразным биологическим фильтром, а сама она представляет собой набор фильтров с перекрывающимися полосами, которые в совокупности покрывают весь набор частот, доступных восприятию человека.

<sup>26</sup> Показателем активности нейрона является частота генерируемых им разрядов, которая тем выше, чем выше амплитуда данной частотной составляющей сигнала.

## **Перцептивная сегментация речевого сигнала**

§ 74. Перцептивная сегментация речевого сигнала является важнейшим этапом восприятия. Она в первую очередь основывается на выделении в сигнале фрагментов, соотносимых со сменой источника звука и связанным с этим изменением степени сужения речевого тракта, что приводит к быстрым изменениям в спектре звуков. Для перцептивной сегментации наиболее полезными являются такие акустические явления, как резкое начало или прекращение звуковых колебаний, свидетельствующее о наличии паузы; резкий переход от интенсивного периодического сигнала к менее интенсивному аperiodическому (шумовому) и наоборот и т.д.

При обнаружении таких событий в сигнале могут быть выделены следующие наиболее яркие акустические фрагменты:

- пауза (участок с нулевой энергией, соответствующий смычке глухого согласного),
- звонкая пауза (энергия только в области основного тона и периодичность, что соответствует смычке звонкого согласного),
- взрыв (шумовой участок с резким начальным нарастанием энергии),
- аperiodический длительный шумовой участок с нерезким началом (соответствующий фрикативному согласному),
- вокализованный шум (участок, совмещающий наличие шумовых составляющих и признаков периодичности – звонкий фрикативный согласный),
- вокальный участок (отсутствие шума, периодичность, энергия в области основного тона, наличие формантных максимумов – гласный или сонорный согласный).

Существенное значение для восприятия (особенно в случае сочетания гласного с сонорным согласным) имеет членение сигнала на основании резких изменений в его амплитуде, т.е. по перепадам интенсивности (так, гласные интенсивнее сонорных согласных). Перепады интенсивности на вокальных участках могут быть поддержаны резкими изменениями F-картины, как это имеет место на границе гласных с носовыми согласными.

## **Перцептивная сегментация и лингвистическое членение речевого сигнала**

§ 75. Перцептивная сегментация сигнала не всегда совпадает с лингвистическим членением на звуки речи. Она отражает, прежде всего, резкие изменения в работе источников звука и в передаточной функции речевого тракта. Такая сегментация в одних случаях может быть слишком мелкой (полученные фрагменты могут быть лишь частями звуковых единиц), а в других ожидаемые границы могут отсутствовать (полученные фрагменты могут соответствовать целым последовательностям звуков).

Следует иметь в виду, что фонологические признаки на фонетическом уровне имеют достаточно сложную структуру – они состоят из целого ряда акустических и перцептивных параметров, так что взаимнооднозначного соответствия между фонологическим признаком и каким-то одним акустическим событием не наблюдается<sup>27</sup>.

Например, во всех языках глухие согласные могут отличаться от звонких 1) отсутствием голоса, 2) большей общей длительностью, 3) большей длительностью послевзрывной фазы, 4) большей интенсивностью взрыва и/или шума, 5) большим значением ЧОТ в начале последующего гласного, 6) меньшей длительностью соседних гласных и другими фонетическими параметрами. При этом в разных языках (и даже в разных позициях в одном языке) на первый план может выходить то один, то другой акустический параметр, например, первый – в любой позиции в СРЛЯ<sup>28</sup>, второй – в конце слова и третий в начале слова в английском языке, четвертый – в нидерландском языке и т.п.

Несомненно, что существует и сложнейшая система взаимной компенсации этих параметров. Например, если носитель русского языка не может принять решение о глухости/звонкости сегмента по наличию голоса, то он может обратиться к значениям других параметров (длительности, интенсивности и т.п.).

## **Акустические ключи звуков русского языка**

§ 76. В целом, акустические ключи практически совпадают с полезными признаками, используемыми при анализе спектрограмм.

Для *гласных* и *сонорных* согласных наиболее важные для опознания ключи находятся на их собственных участках. Для восприятия сонорных звуков существенно взаимное расположение первой, второй и иногда третьей формант по частоте, а также характер изменения этой частоты (направление и скорость формантных переходов). Тип формантных движений зависит как от качества самого гласного, так и от места образования и дополнительной артикуляции соседних согласных (вследствие коартикуляции), а также от просодических условий (например, места ударения и фразового акцента – вследствие артикуляционной редукции). Стационарные участки гласных в обычной речи практически отсутствуют, и для восприятия гласных, а также соседних с ними согласных, формантные переходы оказываются более существенными. При этом наибольшей информативностью (но и наибольшей вариативностью) обладают значения второй форманты, которые отражают положение языка на горизонтальной оси.

<sup>27</sup> Хотя на начальном этапе развития акустической фонетики предполагалось именно такое соотношение фонологических признаков и акустических параметров; эта идеология отражена, например, в классической работе Романа Якобсона, Гуннара Фанта и Морриса Халле "Введение в анализ речи" (см. Новое в лингвистике. Вып. II. М., 1962).

<sup>28</sup> В этом случае соответствующий фонологический признак принято называть "глухость/звонкость", во всех остальных – "напряженность/ненапряженность".

*Носовые* сонорные лучше всего опознаются по переходам гласного, ослаблению  $F_2$  и  $F_3$  в результате антирезонанса и резонансу в области ниже 500 Гц. Очень короткая (около 30 мс.) пауза (одна, реже две или три), окруженная более длительными вокальными фрагментами типа гласных, является признаком *дрожащего* согласного. *Боковой* сонорный [л] характеризуется отчетливой F-картиной, подобной гласным, с низкими локусами  $F_1$  и  $F_2$  (в области 300 и 1000 Гц соответственно) и высокой  $F_3$  в области 2500-3000 Гц. *Аппроксимант* [й] характеризуют минимальная  $F_1$  (ниже 300 Гц)<sup>29</sup> и максимальная для вокальных звуков частота  $F_2$  (2500-3000 Гц). При увеличении степени сужения в спектре [й] проявляются черты, свойственные фрикативным согласным: появляются шумовые составляющие и антирезонансы, эти свойства в русском языке характеризуют фрикативный [й].

Среди *шумных* согласных *взрывные* выделяются по наличию полной смычки (паузы) и взрыва (резкого выброса энергии), а также по быстрым переходам от взрыва к следующим гласным и от предшествующих гласных к смычке, которые возникают вследствие быстрого изменения формы речевого тракта. Эти ключи гораздо надежнее и устойчивее, чем те, по которым определяется место артикуляции.

*Фрикативные* согласные распознаются, в первую очередь, по относительно долгому шумовому периоду с плавным началом. Легко отделяются сибиланты (типа [с] и [ш]) от других фрикативных, у которых нет ярко выраженных пиков энергии и гораздо меньше интенсивность шума.

*Аффрикаты* выделяются по резкому началу относительно долгой шумовой составляющей (так, например, [ц] с отрезанным началом воспринимается как [ц] и соотношению длительностей смычки и шума: у взрывных длительность паузы больше длительности последующего шумового участка, а у аффрикат – меньше).

Для восприятия *места артикуляции* особенно полезными признаками являются 1) область максимума энергии шума относительно формант гласного (рис. 6.14) и 2) характер перехода  $F_2$  от или к соседнему гласному. Если основная область энергии взрыва – более 3000 Гц., то сигнал воспринимается как [т] в контексте любого гласного, а если эта область ниже 500 Гц., то сигнал всегда воспринимается как [п]. В условиях отсутствия взрыва место артикуляции твердого согласного хорошо распознается по переходу  $F_2$  (хотя его направление зависит еще от самого гласного).

*Взрывность* сегмента распознается по наличию голоса (низкочастотного периодического сигнала), а также по меньшей (сравнительно с глухими согласными) интенсивности шума и длительности.

Наиболее информативным ключом для восприятия *мягкости* согласных является локусное значение  $F_2$  в области около 2500 Гц. и соответствующие ему формантные переходы соседних гласных.

## Лингвистический этап восприятия

§ 77. На основе выделенных акустических ключей слушающий строит лингвистическое описание звуковой стороны речевого сообщения, т.е. определяет состав образующих его фонемных единиц и просодических показателей (ударений, элементов интонационного оформления). Эта деятельность относится к последнему, собственно лингвистическому этапу звукового восприятия речи.

В обычных речевых условиях звуковое восприятие подчинено более важной цели: опознанию слов, из которых состоит речевое сообщение. Для этого слушающий располагает гораздо более мощными источниками информации, чем извлеченные из речевого сигнала акустические ключи. Наиболее важный и надежный источник – это знание языка, на котором построено речевое сообщение, то есть

- знание языковых знаков, а значит, и тех фонемных цепочек, которые соответствуют реальным словам (в русском языке нет слова *камán* или *селёный*, а слова *кабáн*, *зелёный* или *солёный* есть);
- знание законов сочетаемости фонем (в начале русского слова невозможна последовательность /нрк/);
- знание законов сочетаемости слов по смыслу (невозможно словосочетание *толстая мысль*);
- знание законов оформления синтаксических связей (в русском языке невозможно словосочетание *стена обвалился*).

Использование этих постоянных языковых знаний, а также результатов текущего восприятия (т.е. совокупности слов или их частей<sup>30</sup>, уже опознанных к данному моменту процесса распознавания) позволяет людям воспринимать и понимать речь в условиях сильных акустических помех и искажений речевого сигнала. Однако фонетическая интерпретация речевого сигнала возможна и без обращения к словарю – например, при звуковом восприятии слогов и бессмысленных звукосочетаний (псевдослов и псевдофраз), а также неизвестных слушающему слов, например, собственных имен.

## Нелинейность и категориальность восприятия

§ 78. Значительная часть информации о сегменте может быть заключена в другом сегменте. Например, информация о месте образования и твердости/мягкости согласных заключена в переходных участках соседних гласных; гласные с переходным участком в начале звучания воспринимаются как сочетание "согласный + гласный" даже при полном отсутствии шума; один и тот же взрыв может восприниматься как [п], [т] или [к] в зависимости от качества следующего гласного и т.п. Поэтому решение о категориальной принадлежности сегмента часто (а в случае естественной речи – почти всегда) слушающий принимает не только на основании звучания самого сегмента, а после анализа следующего звука или целого комплекса (слога или слова), в который входит данный звук. Это свойство восприятия речи человеком называется *нелинейностью*.

<sup>29</sup> Поскольку степень сужения при его произнесении больше, чем при степени сужения при артикуляции [й].

<sup>30</sup> Например, если установлен фонемный состав первого слога, то эти знания могут использоваться для выделения в слове слушающего активного класса слов, которые с этого слога начинаются.

Человек гораздо лучше воспринимает не простые последовательности сегментов, а определенным образом организованные при помощи суперсегментных средств (ударения и интонации). Видимо, при восприятии человек анализирует слоги (в пределах которых действуют правила коартикуляции) и слова (ритмическая структура слова является одним из самых устойчивых элементов сигнала), а решения принимает уже в терминах сегментов<sup>31</sup>. При этом, хотя в идеале человек распознает именно сегменты, часто (когда информации недостаточно) он может оперировать и их признаками (т.е. приписывать распознаваемым единицам значение некоторого класса). Например, если на основании акустического сигнала невозможно принять решение о глухости/звонкости распознаваемого сегмента, принятие решения может быть отложено; в этом случае оно будет принято на основании знания словаря или контекста.

**§ 79.** Если предъявить носителю русского языка несколько разных искусственно созданных согласных, отличающихся только длительностью звонкого участка смычки, и попросить его определить, какой согласный он слышит в каждом конкретном случае, то каково бы ни было число градаций длительности звонкого отрезка согласного, носитель русского языка будет относить его лишь к одной из двух категорий – глухой или звонкой.

Различные изменения направления движения  $F_2$  соседнего гласного со стационарной частью в области 1200 Гц. ведут к восприятию всего трех категорий согласных звуков независимо от числа стимулов: [п] – если  $F_2$  резко восходящая, [к] – если она плавно восходящая, ровная или плавно нисходящая и [т] – если резко нисходящая. Таким образом, в восприятии не существует плавных переходов, вместо этого происходит резкое смещение из одной категории в другую. Эта особенность восприятия речи человеком называется *категориальностью* восприятия. В неречевых сигналах ситуация совсем другая – там наблюдаются градуальные изменения в восприятии.

Категориальность восприятия позволяет гораздо быстрее анализировать речевой сигнал – в секунду человек способен воспринимать до 30 речевых единиц и только одну неречевую.

### **Фонемная интерпретация сообщения (основные модели восприятия речи)**

**§ 80.** Задача фонемной интерпретации речевого сообщения состоит в том, чтобы преобразовать слуховое изображение речевого сигнала, в котором выделены фонетически полезные акустические события и признаки, в последовательность полностью или частично опознанных фонемных единиц.

Один из важнейших вопросов перцептивной фонетики – это вопрос о том, каким именно образом извлекается информация из акустического сигнала. Существует несколько гипотез относительно способности человека "узнавать" или распознавать в речевом сигнале последовательности фонем, из которых состоят языковые знаки в памяти человека; эти гипотезы можно разделить на три класса:

- гипотеза артикуляторного источника, наиболее известным вариантом которой является моторная теория речевосприятия (фонологический анализ посредством "внутреннего" синтеза);
- гипотеза перцептивного эталона, согласно которой основным принципом считается прямая идентификация (сравнение с имеющимся в памяти языка набором фонем, представленных в удобном для распознавания виде);
- признаковая гипотеза (распознавание на основе логических правил выбора возможных фонемных решений).

**§ 81.** Согласно *моторной теории восприятия речи* предполагается, что в процессе восприятия слушающий на основе анализа акустических признаков восстанавливает целевые артикуляции, осуществление которых привело к образованию сигнала с наблюдаемыми акустическими характеристиками. Различные акустически сигналы, возникающие при произнесении одного и того же звукотипа в разных контекстных условиях (например, гласный [а] в слове *пат* и [ä] в слове *пять*) воспринимаются одинаково не потому, что обладают общими акустическими признаками, а потому, что при их производстве используются одни и те же комплексы артикуляционных движений. Таким образом, согласно этой теории, слушающий при восприятии создает модель артикуляции, которая позволила бы произвести услышанный им сигнал.

*Гипотеза перцептивного эталона* предполагает, что в памяти носителя языка каждой фонеме соответствует свой перцептивный эталон. Для каждого акустического признака в составе эталона задаются

<sup>31</sup> "Обращение к словам как оперативным единицам восприятия речи объясняется, по крайней мере, тремя взаимосвязанными причинами. Во-первых, для пофонемного восприятия недостаточна разрешающая способность слухового анализатора человека: при необходимости постоянно осуществлять текущее перекодирование поступающего акустического сигнала в цепочку фонем слуховой анализатор должен перерабатывать слишком большое количество информации в единицу времени. Во-вторых, значительная часть акустических сегментов "нормальной" речи характеризуется неполным типом произнесения; их объективных характеристик просто недостаточно для того, чтобы дать им фонемную интерпретацию. Наконец, в-третьих, человеку вообще свойственна тенденция к укрупнению единиц и признаков в процессах восприятия, т.к. это повышает быстроту действия соответствующих механизмов. В указанных условиях слушающий, пользуясь признаками слова как целостной единицы, относит его к некоторому классу. В информационно насыщенной среде, обычной для речевого общения, идентификации слова с точностью до некоторого класса обычно достаточно: к тому моменту, когда появляется необходимость в распознавании данного слова, возможности выбора уже значительно сужены предтекстом, ситуацией, опытом и установкой слушающего." (Зиндер Л.Р., Касевич В.Б. Фонема и ее место в системе языка и речевой деятельности // Вопросы языкознания. 1989, N 6).

допустимые пределы варьирования. Иногда допускается существование для фонемы не одного, а нескольких перцептивных эталонов. В качестве самостоятельных эталонов в этом случае выступают контекстные реализации фонем, которым соответствуют разные комплексы акустических признаков, хотя и признается, что в таком случае набор перцептивных эталонов будет весьма большим.

*Признаковая гипотеза* предполагает, что фонема существует в памяти носителя языка в виде набора значений фонологических различительных признаков и не имеет прямого перцептивного коррелята. Распознавание фонемы осуществляется через распознавание ее фонологических признаков.

Ни одна из изложенных выше гипотез не достигла той степени проработки и детализации, чтобы можно было сравнивать их с точки зрения соответствия реальному поведению человека. Кроме того, человеческий мозг устроен так сложно и возможности его так велики, что реальное речевое поведение человека может базироваться на элементах всех трех гипотез, а, может быть, и на каких-то еще возможностях, которые пока что не учтены.

## Методы изучения восприятия

§ 82. Для изучения восприятия требуются особые методы, так как в реальных процессах коммуникации очень многое опирается на контекст, а сам процесс восприятия сильно автоматизирован – для того, чтобы понять, как принимается решение, например, о твердости или мягкости услышанного согласного, нельзя просто спросить носителя языка: "Как Вы опознаете мягкость?"

При экспериментальном исследовании восприятия необходимо определить тип речевого материала, способ его получения и реакцию испытуемых. В качестве материала перцептивного исследования могут служить звуки, фрагменты меньшие, чем звук, слоги, слова, фразы, а также иные звуковые последовательности. При этом следует иметь в виду, что осмысленные речевые отрезки (слова или фразы) мало пригодны для выявления фонетически полезных признаков речевого сигнала, так как при их восприятии результат зависит от подсказки, которая создается знанием словаря, действующими в языке правилами синтаксической и смысловой сочетаемости слов. Поэтому в перцептивных экспериментах чаще всего используются псевдослова или псевдофразы – логотомы (бессмысленные звуковые комплексы, которые построены в соответствии со звуковыми законами данного языка).

Материалом эксперимента могут быть отрезки естественной речи (выделенные из произнесенных слов) или искусственной (синтезированной) речи, поскольку при анализе фонетически полезных признаков необходимо иметь возможность систематически изменять акустические параметры с тем, чтобы проверить, как эти изменения влияют на восприятие.

В экспериментах по восприятию могут анализироваться разные типы реакций испытуемых – например, опознание и различение. В опытах на опознание испытуемый должен идентифицировать звуковой стимул, обозначив его некоторым условным образом, например, буквами алфавита или транскрипционными символами (при этом может разрешаться и отказ от ответа). Предполагается, что в опытах на опознание слушающий пытается тем или иным образом отождествить воспринимаемый сигнал с имеющимися у него внутренними звуковыми образами (фонемами или их контекстными реализациями). В экспериментах на различение испытуемым предъявляются два отдельных звуковых стимула, и он должен ответить, отличаются они друг от друга или нет (эти опыты имеют меньшую ценность, так как человеческая способность к различению очень велика).

## ЛИТЕРАТУРА.

*Зиндер Л.Р.* Общая фонетика. М., 1979.

*Кодзасов С.В., Кривнова О.Ф.* Общая фонетика. М., 2001.

*Речь. Артикуляция и восприятие.* Пред. Л.А.Чистович и В.А. Кожевникова. М.-Л., 1965.

*Джапаридзе З.* Перцептивная фонетика. Тбилиси, 1985.

*Зиновьева Н.В.* Система акустических ключей к распознаванию фонетических единиц русского языка // Экспериментальная фонетика. М., 1989.

*Физиология речи. Восприятие речи человеком.* Пред. Л.А.Чистович и А.В.Венцова. Л., 1976.

*Фланаган, Дж.* Анализ, синтез и восприятие речи. М, 1968.