

Основные спектральные характеристики речевых сигналов и их определение

С. А. Иванушкин, email: nr.kvvu@yandex.ru ¹

М. И. Поддубный, email: podd.maxim@yandex.ru ²

Д. А. Косарев ², А. В. Вешкин ², П. С. Заборских ², Ю. И. Елгин ²

¹ КВВУ им. генерала армии С. М. Штеменко

² ФГАУ ВИТ «ЭРА»

***Аннотация.** В работе представлен перечень основных частотных, временных и энергетических характеристик речевого сигнала и их определение в программной среде.*

***Ключевые слова:** речевые сигналы, спектральные характеристики, частотные характеристики, временные характеристики, энергетические характеристики, речеподобная помеха, маскирование речевого сигнала.*

Введение

В системах активной виброакустической защиты речевой информации используются, главным образом, защитные помехи типа «Белый шум» или его модификации [2]. Однако в последнее время более пристальное внимание привлекает речеподобная помеха. Применение речеподобной помехи позволяет в ряде случаев снизить требуемые уровни защитного сигнала и уменьшить эффективность средств шумоочистки.

Еще в середине 50-х годов прошлого века отечественные исследователи Л.А. Варшавский и И.М. Литвак высказали гипотезу о том, что акустическое качество звуков речи определяется соотношением уровней сигнала в полосах спектра. Форманты являются лишь доступным для речевого аппарата человека способом достижения необходимых полосных соотношений.

При расширении исследований в рамках задачи формирования речеподобных шумовых помех возникают новые вопросы. Это требует развития теории определения акустического качества звуков речи применительно к речевому сигналу, то есть определению тех или иных спектральных характеристик говорящего по параметрам его голоса.

Таким образом, существует необходимость в определении основных спектральных характеристик речевых сигналов с целью их

последующего применения при формировании речеподобной помехи для ее использования при ведении защищаемых конфиденциальных переговоров [4].

В работе приведен перечень частотных, энергетических и временных спектральных характеристик, и разработана методика их определения.

1. Перечень основных спектральных характеристик речевых сигналов

Спектральные характеристики всех звуковых сигналов в общем случае подобны и описаны в большинстве книг курса общей физики. Определяются они с помощью двух основных зависимостей, раскрывающих их состав в полной мере: зависимость амплитуды от времени (осциллограмма или амплитудно-временная характеристика) и зависимость амплитуды от частоты (спектрограмма или амплитудно-частотная характеристика). Особенностью же речевых сигналов является наличие формантных областей – набора гармоник, образующих определенный звук речи [7].

Все речевые сигналы можно условно подразделить на то, что схоже у определенного состава людей – это отношение к какой-либо языковой группе: подобный набор произносимых слов и звуков, и то, что индивидуально для каждого человека – ряд уникальных особенностей, выраженных в физической индивидуальности речи, произношении, тембра голоса, так называемом артикуляционном аппарате человека. Данные особенности позволяют идентифицировать человека. Для формирования наиболее эффективной речеподобной помехи наиболее важными параметрами являются именно индивидуальные особенности речи [3].

При формировании перечня спектральных характеристик из формантных областей берется в расчет только первая (основная) гармоника, так как она является наиболее информативной. Остальные гармоники лишь определяют произнесенный звук, что в большинстве случаев является общим для одного языкового состава людей.

В соответствии с измеряемой величиной основные спектральные характеристики разделены на следующие подгруппы: частотные, энергетические и временные. Речевой сигнал является не постоянной величиной и для полной картины его состава необходим контроль значений в каждый момент времени, поэтому для каждой характеристики определяется ее минимальное, среднее и максимальное значение.

Важным уточнением является то, что речевой сигнал является непрерывным во времени, и под гармоникой понимается некоторая спектральная плотность с выраженным пиком амплитуды.

Частотные характеристики речи:

1. частота основного тона – F_{OT} (Гц) – частота первой гармоники спектра;
2. период основного тона – $T_{OT} = 1 / F_{OT}$ (мс);
3. количество побочных гармоник (формантных областей) – N_{Γ} ;
4. нижняя частота спектра – $f_H = F_{OT}$ (Гц);
5. верхняя частота спектра – f_B (Гц);
6. частотный диапазон – $\Delta f = f_B - f_H$ (Гц);
7. частота максимального уровня спектральной плотности – F_{max} (Гц).

Энергетические характеристики речи:

1. нижний уровень громкости речи – I_H (дБ);
2. верхний уровень громкости речи – I_B (дБ);
3. динамический диапазон – $\Delta I = I_B - I_H$ (дБ);
4. амплитуда основного тона – I_{OT} (дБ).

Временные характеристики речи:

1. длительность звука речи – $\tau_{зв}$ (мс);
2. длительность пауз речи между словами – $\tau_{пс}$ (мс);
3. длительность пауз речи между фразами – $\tau_{пф}$ (мс);
4. скорость звуков речи – $v_{зв}$ (звуков/с);
5. темп речи – $v_{сл}$ (слов/мин);
6. плотность речи – ρ_r (%) – отношение времени наличия звука к полному времени речевого сигнала;
7. скорость изменения уровня громкости основного тона – $v_{гот}$ (дБ/с);
8. скорость изменения частоты основного тона – $v_{чот}$ (Гц/с).

2. Определение основных спектральных характеристик речевых сигналов

В соответствии с перечнем основных характеристик представлена методика их определения, с помощью аудио редактора и плагина анализатора спектра звуковых сигналов.

Перед тем как приступить к исследованию основных характеристик речевых сигналов необходимо подготовить звуковые файлы: выровнять их средние значения программной громкости, длительность файлов, определить начальные настройки звуковых плагинов. Для выравнивания среднего значения программной громкости необходимо использовать линейные усилители, которые не вносят дополнительных искажений в исследуемые сигналы.

В цифровом представлении сигналов речь идет именно о программной громкости, характеризующей насколько громко слушателем, воспринимается материал. Данная величина является относительной и рассчитывается по специальному алгоритму, учитывающему восприятие звука человеком, разработанному ITU - BS.1770 [6, 8, 9].

Порядок определения спектральных характеристик представлен для одного звукового файла. Некоторые характеристики объединены в один пункт из-за схожего метода определения значений.

3. Определение частотных характеристик

Частота основного тона (ЧОТ). Количество побочных гармоник (ПГ). Нижняя частота спектра (НЧС). Верхняя частота спектра (ВЧС). Частота максимального уровня спектральной плотности (ЧМУСП)

Для определения данных характеристик необходимо использовать spectrogramму. Порядок определения значений частотных характеристик:

1. Зациклить и воспроизвести временной промежуток с выраженным участком речи, примерной длительностью 0.5-1 секунды.
2. Зафиксировать спектр зацикленного участка звуковой записи.
3. С помощью курсора и индикаторов определить значения:
 - ЧОТ = НЧС – значение частоты основного тона;
 - ПГ – количество гармоник в спектре;
 - ВЧС – значение верхней частоты в спектре сигнала;
 - ЧМУСП – значение частоты с максимальным значением амплитуды в спектре сигнала.
4. Записать полученные значения в экспериментальные таблицы.
5. Повторить п. 1-5 со следующим временным промежутком до конца звукового файла.

Период основного тона

Период основного тона определяется с помощью частоты основного тона по формуле (1):

$$T_{от} = 1 / F_{от} [с] \quad (1)$$

где $F_{от}$ – частота основного тона в Гц.

Значения заносятся в экспериментальную таблицу.

Частотный диапазон

Частотный диапазон определяется с помощью НЧС и ВЧС по формуле (2):

$$\Delta f = f_B - f_H [Гц] \quad (2)$$

где f_B – верхняя частота в спектре в Гц, f_H – нижняя частота в спектре в Гц.

4. Определение энергетических характеристик

Нижний уровень громкости речи (НУГР). Верхний уровень громкости речи (ВУГР). Амплитуда основного тона (АОТ)

Для определения данных характеристик необходимо использовать спектрограмму. Порядок определения значений энергетических характеристик:

1. Зациклить и воспроизвести временной промежуток с выраженным участком речи, примерной длительностью 0.5-1 секунды.
2. Зафиксировать спектр зацикленного участка звуковой записи.
3. С помощью курсора и индикаторов определить значения:
 - НУГР – уровень громкости, соответствующий амплитуде гармоники с наименьшей амплитудой;
 - ВУГР – уровень громкости, соответствующий амплитуде гармоники с наибольшей амплитудой;
 - АОТ – уровень громкости, соответствующий амплитуде первой гармоники спектра.
4. Записать полученные значения в экспериментальные таблицы.
5. Повторить п. 1-5 со следующим временным промежутком до конца звукового файла.

Динамический диапазон

Динамический диапазон определяется с помощью полученных значений в предыдущем пункте. Значения динамического диапазона рассчитываются по формуле (3):

$$\Delta I = I_B - I_H [дБ] \quad (3)$$

где I_B – верхний уровень громкости речи в дБ, I_H – нижний уровень громкости речи в дБ.

Полученные значения заносятся в экспериментальную таблицу.

5. Определение временных характеристик

При определении временных характеристик не принимается в расчет физиологическая и акустическая классификация звуков речи [1]. Под звуком речи понимается минимальная произносимая единица языка, обладающая определённым спектральным составом.

Длительность звука речи (ДЗР)

Длительность звука речи определяется с помощью временной шкалы в аудио редакторе. Порядок определения значений длительностей звука речи:

1. Зациклить и воспроизвести временной промежуток с выраженным участком речи, примерной длительностью 100 мс (длительность одного звука в среднем составляет 20-350 мс [5]).
2. С помощью передвижения границ цикла добиться звучания одного звука.
3. Определить значения времени начала и конца цикла и вычислить длительность звучания одного звука.
4. Записать полученное значение в экспериментальную таблицу.
5. Повторить п. 1-5 со следующим временным промежутком до конца звукового файла.

Длительность паузы между словами (ДПМС)

Длительность паузы речи также определяется с помощью временной шкалы в аудио редакторе. Порядок определения значений длительностей паузы речи:

1. Зациклить и воспроизвести временной промежуток с выраженным участком отсутствия звука речи (участок следует выбирать между произнесенными подряд словами).
2. С помощью передвижения границ цикла добиться полного отсутствия звучания звука речи.
3. Определить значения времени начала и конца цикла и вычислить длительность паузы речи.
4. Записать полученное значение в экспериментальную таблицу.
5. Повторить п. 1-5 со следующим временным промежутком до конца звукового файла.

Длительность паузы между фразами (ДПМФ)

Определения значений длительностей паузы речи между фразами идентичен определению значений длительностей пауз речи между словами. Различие заключается в выборе участка отсутствия звука речи. При определении ДПМФ участок выбирается между законченными

фразами, данный участок, как правило, продолжительнее участка ДПМС.

Скорость звуков речи (СкЗР)

Единицей измерения скорости речи является количество звуков, произнесенных в секунду, и определяется с помощью временной шкалы в аудио редакторе. Порядок определения значений скорости речи:

1. Воспроизвести временной промежуток, с выраженным участком речи, длительностью 1 с.
2. Посчитать количество произнесенных звуков.
3. Записать полученное значение в экспериментальную таблицу.
4. Повторить п. 1-4 со следующим временным промежутком до конца звукового файла.

Темп речи (ТР)

Единицей измерения темпа речи является количество слов, произнесенных в минуту. Для его определения необходимо посчитать количество произнесенных слов в речевом сигнале и привести полученный результат к единице измерения – слов/мин по формуле (4):

$$v_{сл} = \frac{60 \cdot N_{сл}}{t_{зв}} \text{ [слов/мин]} \quad (4)$$

где $N_{сл}$ – количество слов в речевом сигнале, $t_{зв}$ – длительность речевого сигнала в секундах.

Записать полученное значение в экспериментальную таблицу.

Плотность речи (ПР)

Плотность речи – это отношение времени наличия звука к полному времени речевого сигнала. Данная величина показывает в процентном соотношении сколько времени из всего сигнала занимает речь. Для определения данной величины необходимо посчитать суммарную длительность всех звуков или пауз. Плотность речи определяется по формуле (5):

$$\rho_p = \frac{\sum_{i=1}^m \tau_{зв i}}{t_{зв}} \cdot 100\% = \frac{t_{зв} - \sum_{i=1}^m \tau_{п i}}{t_{зв}} \cdot 100\% \quad (5)$$

где $\tau_{зв}$ – длительность звука речи в секундах, $\tau_{п}$ – длительность отсутствия звука речи (паузы) в секундах, $t_{зв}$ – длительность речевого сигнала в секундах.

Скорость изменения уровня громкости основного тона (ГОТ).
Скорость изменения частоты основного тона (ЧОТ)

Для определения скорости изменения уровня громкости и частоты основного тона необходимо использовать спектр речевого сигнала. Порядок определения значений:

1. Зациклить и воспроизвести временной промежуток с выраженным участком речи, длительностью 1 с.
2. Зафиксировать спектр зацикленного участка звуковой записи, соответствующий минимуму значений ГОТ и ЧОТ.
3. Определить с помощью курсора и индикаторов:
 - ГОТ – уровень интенсивности, соответствующий амплитуде основного тона.
 - ЧОТ – значение частоты основного тона.
4. Зафиксировать спектр зацикленного участка звуковой записи, соответствующий максимуму значений ГОТ и ЧОТ.
5. Определить с помощью курсора и индикаторов:
 - ГОТ – уровень интенсивности, соответствующий амплитуде основного тона.
 - ЧОТ – значение частоты основного тона.
6. Рассчитать значения скорости изменения ГОТ и ЧОТ по формулам (6, 7):

$$v_{ГОТ} = I_{max} - I_{min} \text{ [дБ/с]} \quad (6)$$

$$v_{ЧОТ} = f_{max} - f_{min} \text{ [Гц/с]} \quad (7)$$

где I_{max} , I_{min} – максимальный и минимальный уровень громкости основного тона в дБ, f_{max} , f_{min} – максимальное и минимальное значение частоты основного тона в Гц.

7. Записать полученные значения в экспериментальную таблицу.
8. Повторить п. 1-8 со следующим временным промежутком до конца звукового файла

Вывод

Результатом работы является перечень основных спектральных характеристик и методика их определения с помощью аудио редактора и плагинов анализатора спектра. Применение указанной методики

позволяет проводить испытания по определению и анализу частотных, временных и энергетических характеристик речевых сигналов. Знание основных спектральных характеристик способствуют более эффективному формированию речеподобной помехи для маскирования речевых сигналов.

Список литературы

1. Бодуэн де Куртенэ, И. А. Введение в языкознание / И. А. Бодуэн де Куртенэ, А. А. Реформатский. – М., 1977.
2. Дидковский, В. С. Акустическая экспертиза каналов речевой коммуникации / В. С. Дидковский, М. В. Дидковская, А. Н. Продеус. – Киев, 2008. – 420 с.
3. Киреев, К. К. Основные параметры и разновидности речеподобных сигналов, применяемых для защиты речевой информации / К. К. Киреев, Н. А. Русин, П. С. Заборских, Ю. И. Елгин // Сборник КВВУ им. Штеменко. – Краснодар, 2020.
4. Поддубный, М. И. Система акустических и виброакустических помех для организации защиты помещений от утечки речевой информации / М. И. Поддубный, К. В. Киреев, Н. А. Русин, П. С. Заборских, Ю. И. Елгин // Сборник статей конференции «Информатика вычислительная техника». – Анапа: ФГАУ ВИТ «ЭРА», 2020.
5. Покровский, Н. Б. Расчет и измерение разборчивости речи / Н. Б. Покровский. – М.: Связьиздат, 1962.
6. Приказ Минкомсвязи России №171. Об утверждении Рекомендаций в области нормирования звуковых сигналов в телерадиовещании – Утв. 2005-05-21.
7. Сапожков, М. А. Акустика : Справочник / М. А. Сапожков. – М. : Радио и связь, 1989.
8. EBU Recommendation R 128. Loudness normalisation and permitted maximum level of audio signals – Введ. авг. 2020. – Женева : Европейский вещательный союз.
9. ITU-R BS.1770. Algorithms to measure audio programme loudness and true-peak audio level – Введ. окт. 2015. – Женева : Международный союз электросвязи.