



Deep Learning School





Yermekova Assel

Skoltech '22
MLE in Noah Ark Lab, Huawei

Lecture 1. Audio representation: wav

Lecture: Intro to Audio

Content

Lecture: Intro to Speech Processing

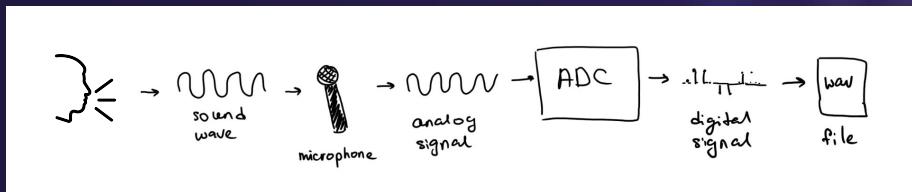


Sound representation:

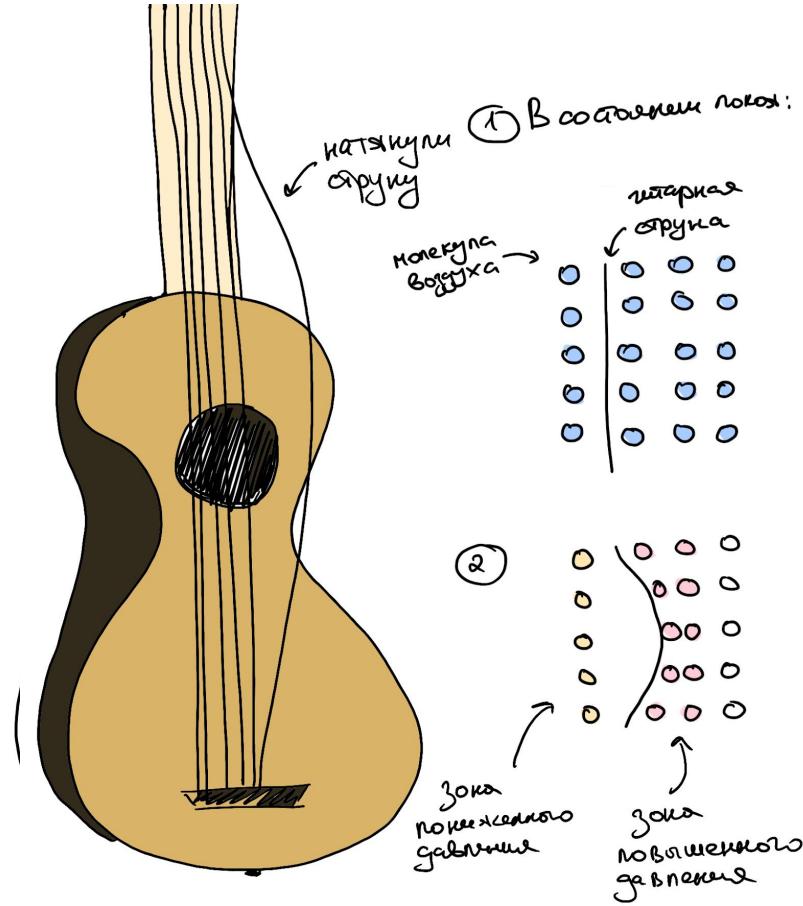
- What is a sound?
- Properties of a sound wave
 - Amplitude
 - Frequency
 - Phase
 - Power, Energy, Intensity, Pressure level
- Discretization
 - in Time (Sample rate, Shannon-Nyquist Theorem, Aliasing)
 - in Amplitude (Bit depth)
 - Number of channels
 - mono, stereo
 - spatial audio
- Audio formats
 - Uncompressed:
 - wav, aiff
 - Lossless compression:
 - flac, alac...
 - Lossy compression:
 - mp3, opus
- Audio plots and its interpretation

What is a sound?

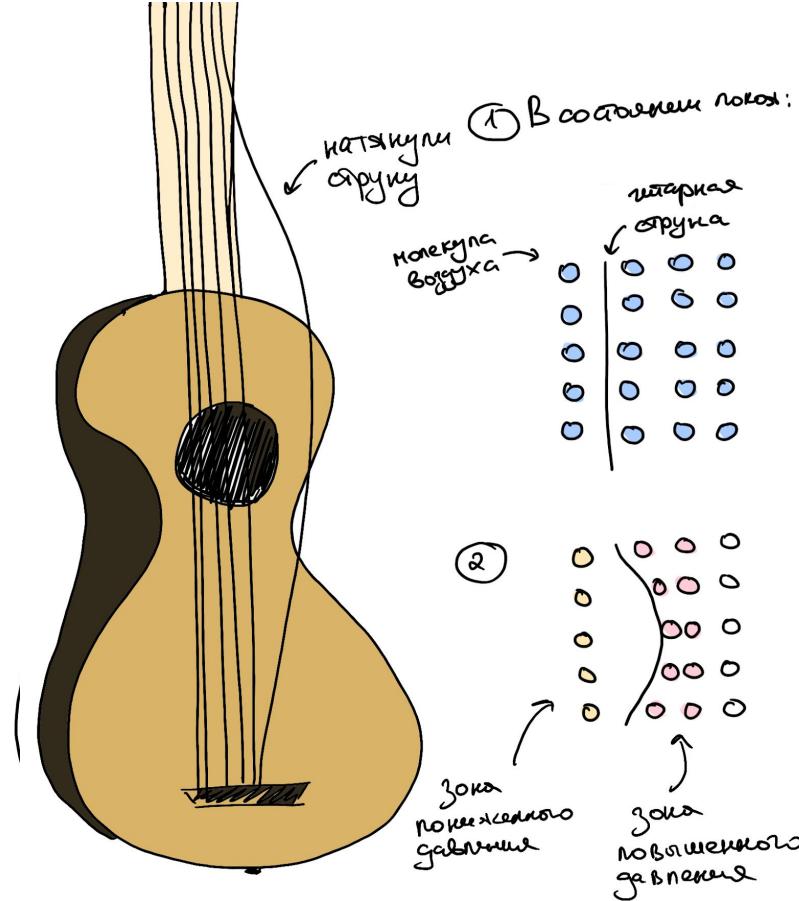
Lecture: Intro to Speech Processing



What is a sound?



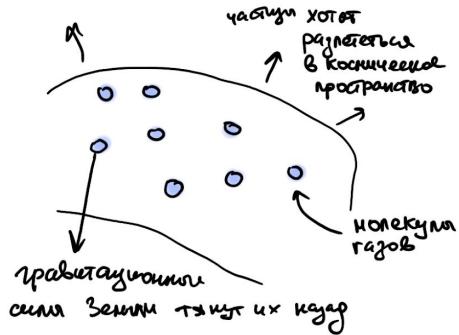
What is a sound?



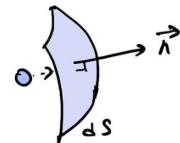
В состоянии покоя у нас существует атмосферное давление.

What is a sound?

В состоянии покоя у нас существует атмосферное давление.



$$p = \frac{dF_n}{dS}$$

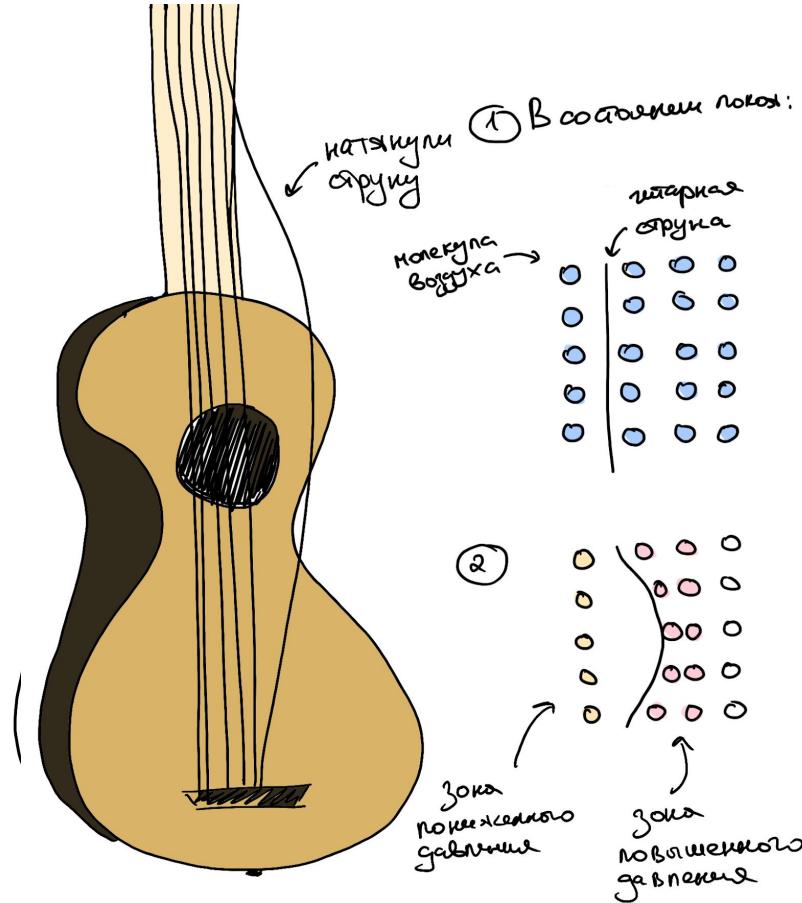


Давление - сила, которую газ оказывает на единицу площади поверхности.

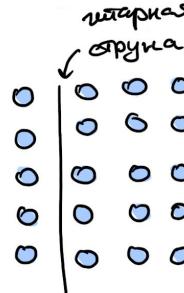
Давление возникает из-за столкновений молекул газа с друг с другом.

Каждое такое столкновение создает крошечную силу. Сумма миллиардов этих столкновений в секунду создает постоянное атмосферное давление.

What is a sound?

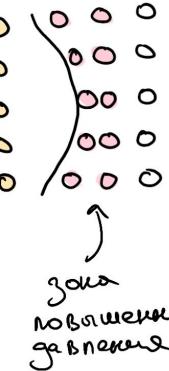


① В состоянии покоя:

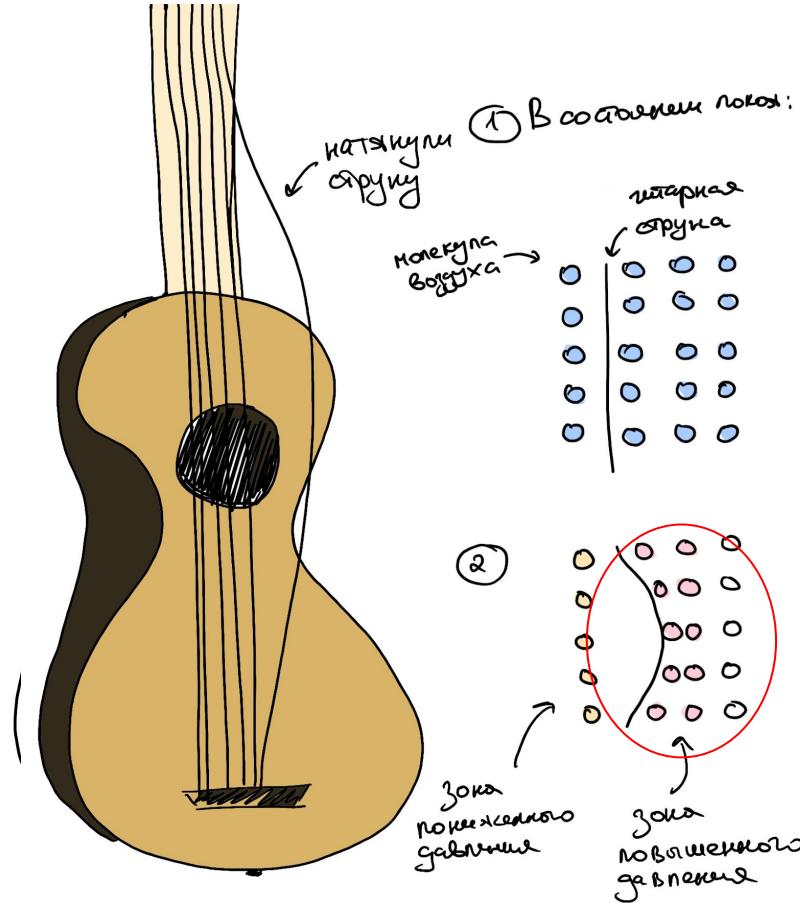


В состоянии покоя у нас существует атмосферное давление.

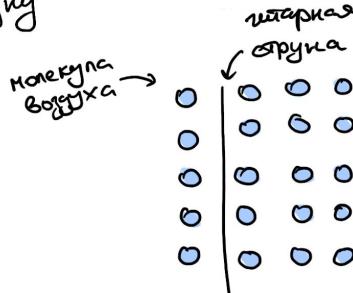
②



What is a sound?



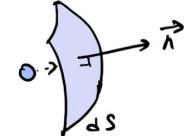
① В состоянии покоя:



В состоянии покоя у нас существует атмосферное давление.

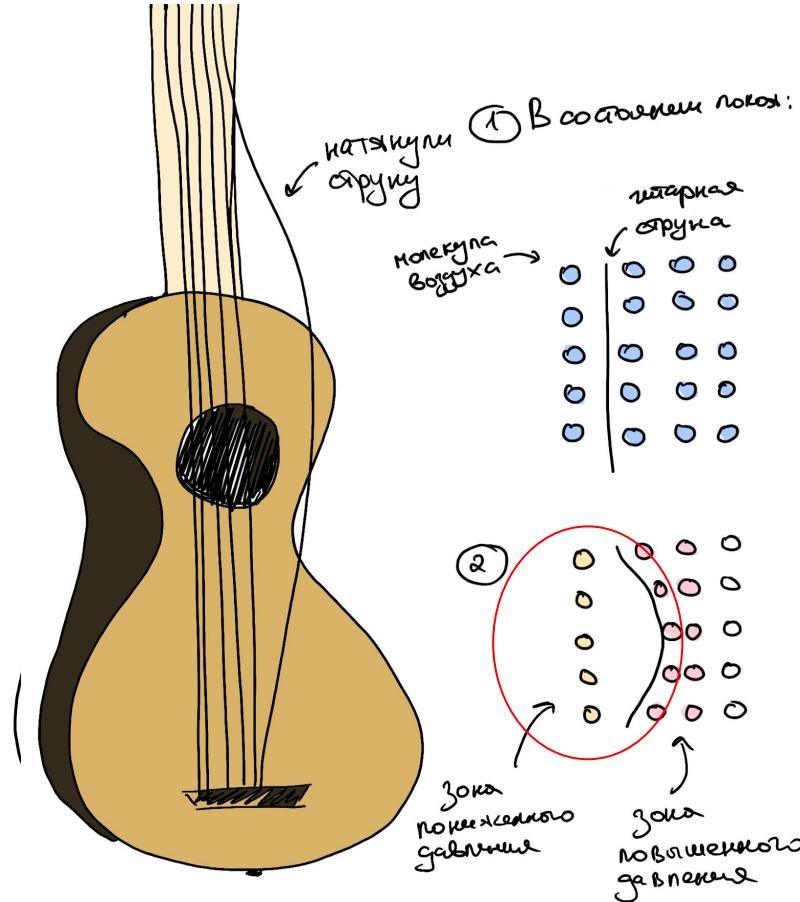
Почему увеличение количества молекул создает область повышенного давления?

$$p = \frac{dF_n}{dS}$$



Давление - сила, которую газ оказывает на единицу площади поверхности.

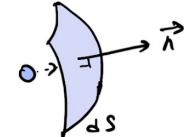
What is a sound?



В состоянии покоя у нас существует атмосферное давление.

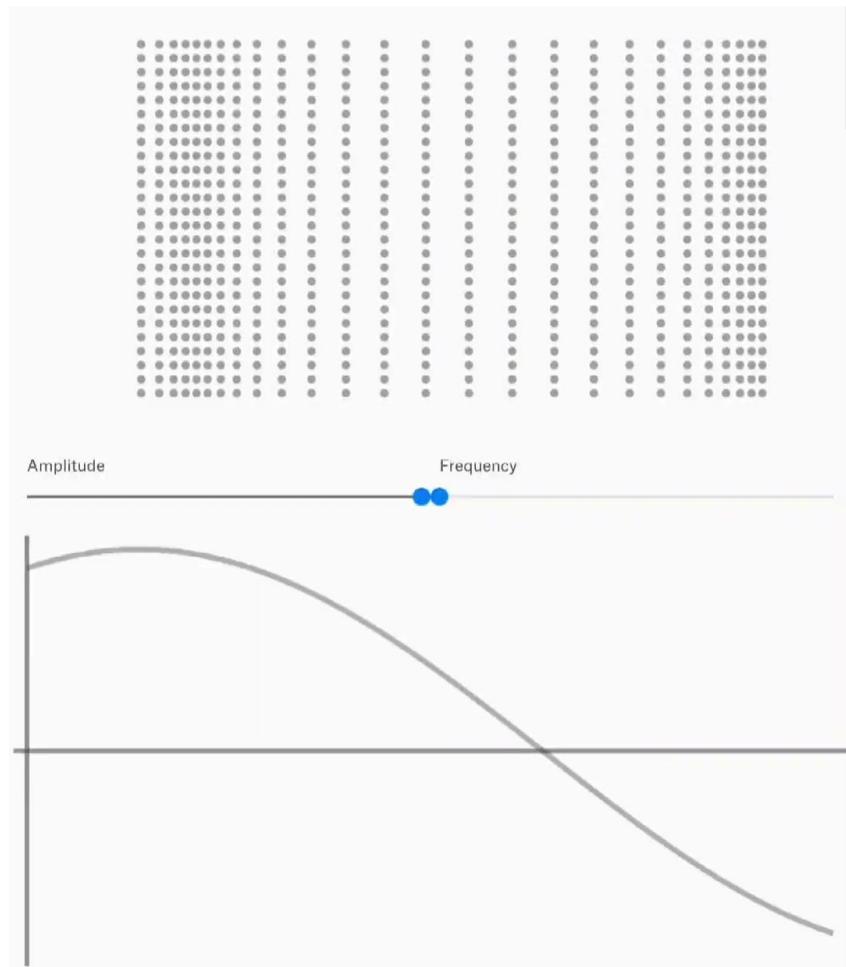
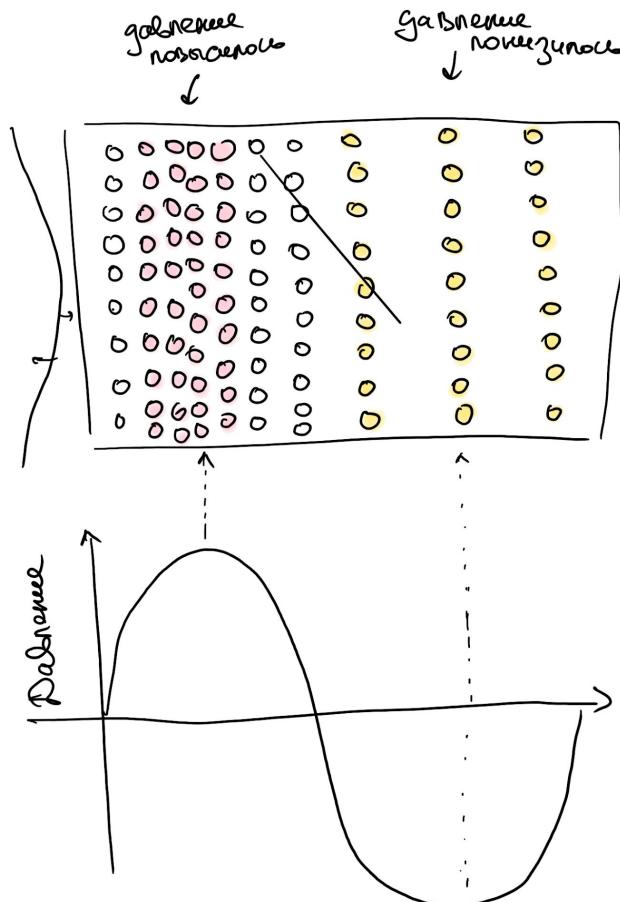
Почему уменьшение количества молекул создает область пониженного давления?

$$p = \frac{dF_n}{dS}$$



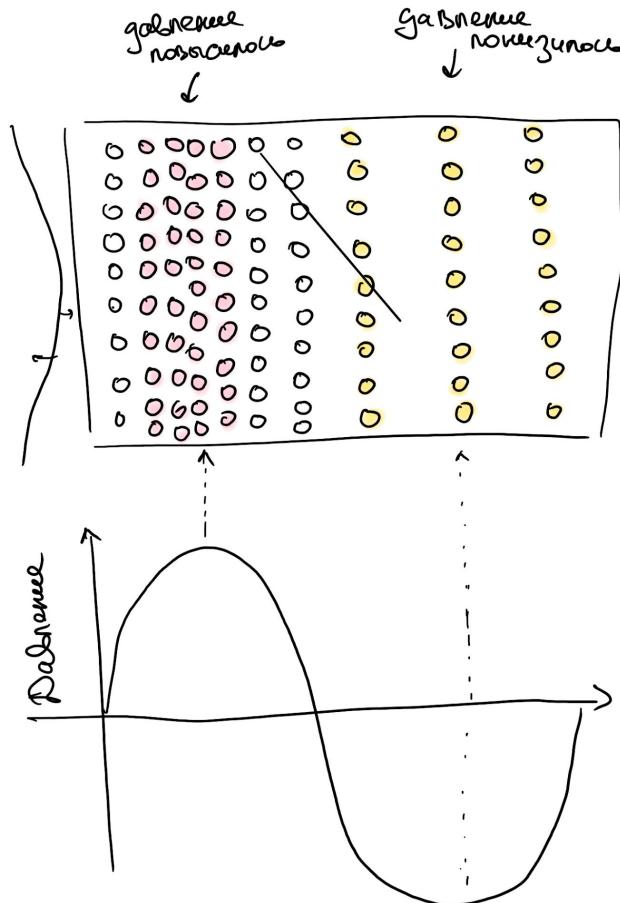
Давление - сила, которую газ оказывает на единицу площади поверхности.

What is a sound?



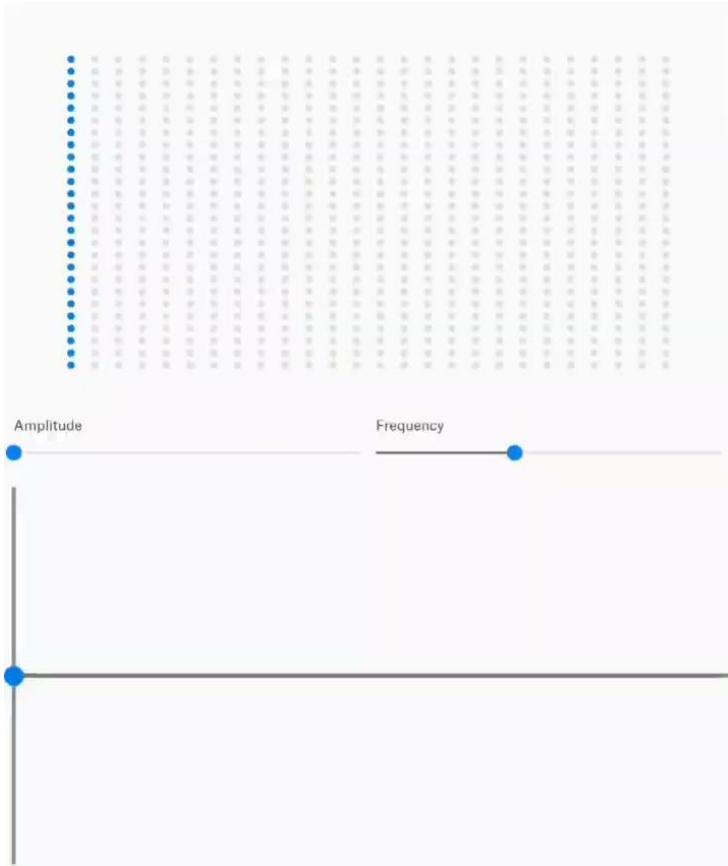
<https://pudding.cool/2018/02/waveforms/>

What is a sound?



Таким образом, звуковая волна это изменение атмосферного давления в воздухе за счет колебаний молекул.

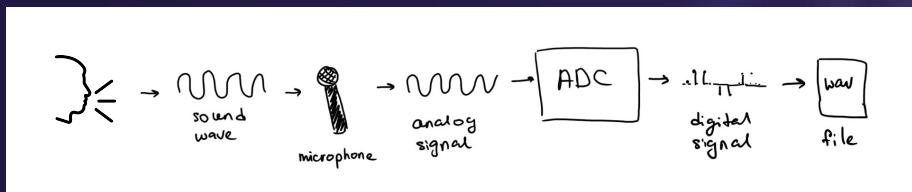
What is a sound?



Важно понимать, что сами молекулы остаются на своих позициях, относительно которых они просто колеблются.

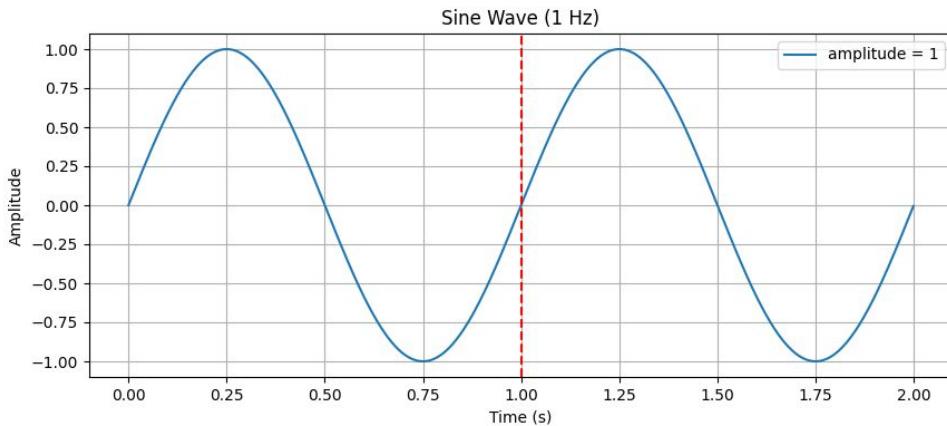
Properties of sound waves

Lecture: Intro to Speech Processing



Properties of Sound Waves: Amplitude

Амплитуда



Давайте все разберем на простом примере простой синусоидальной волны

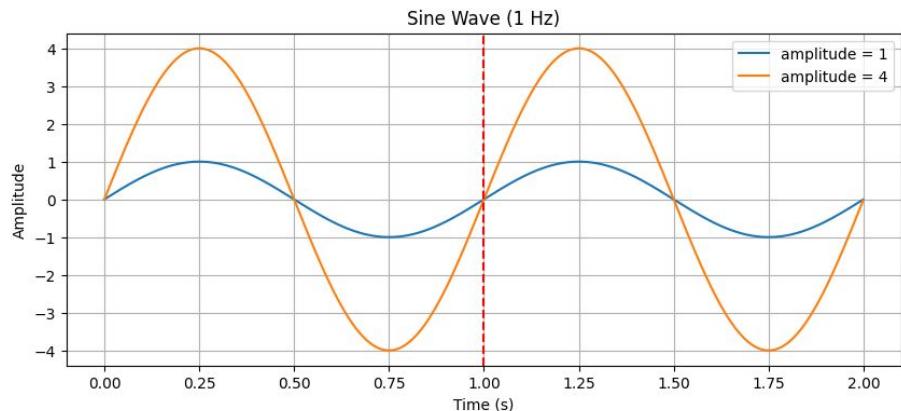
$$y = A \sin(2\pi\omega t)$$

А - амплитуда

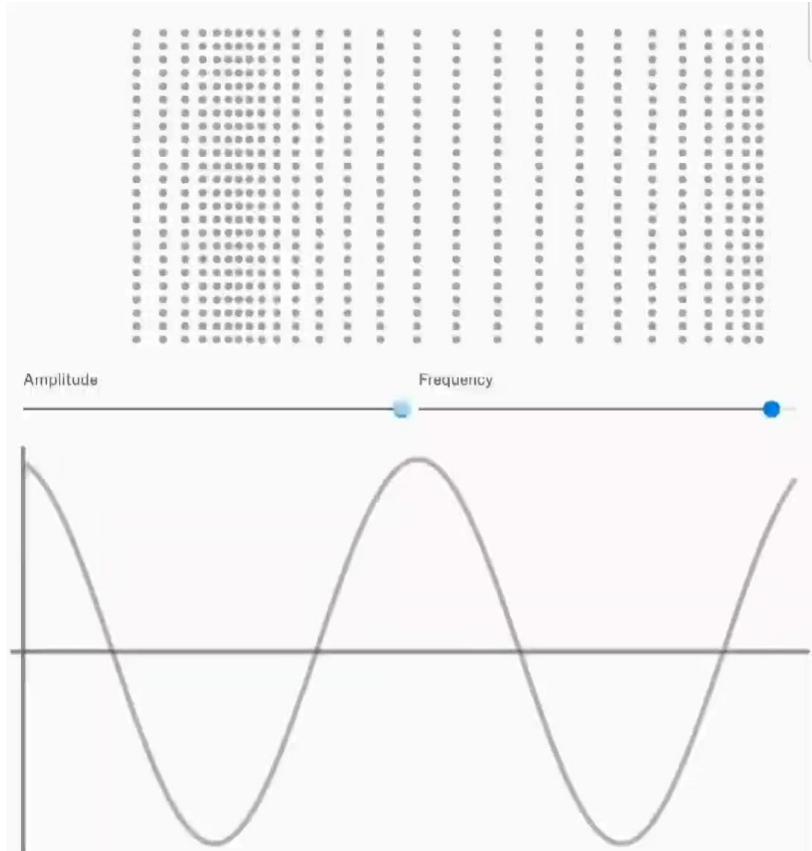
Амплитуда это максимальное отклонение величины от ее значения в равновесном состоянии.

Properties of Sound Waves: Amplitude

Амплитуда



Мы воспринимаем амплитуду как громкость.

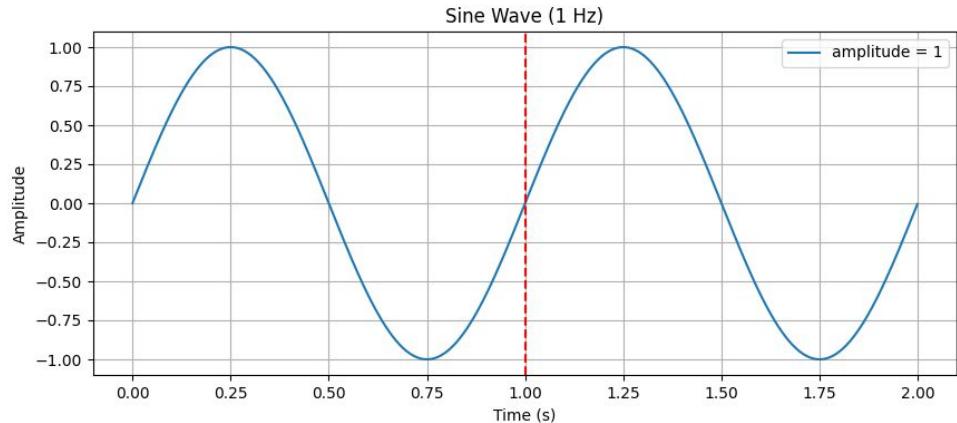


<https://pudding.cool/2018/02/waveforms/>

Properties of Sound Waves: Decibel

Амплитуда [?]

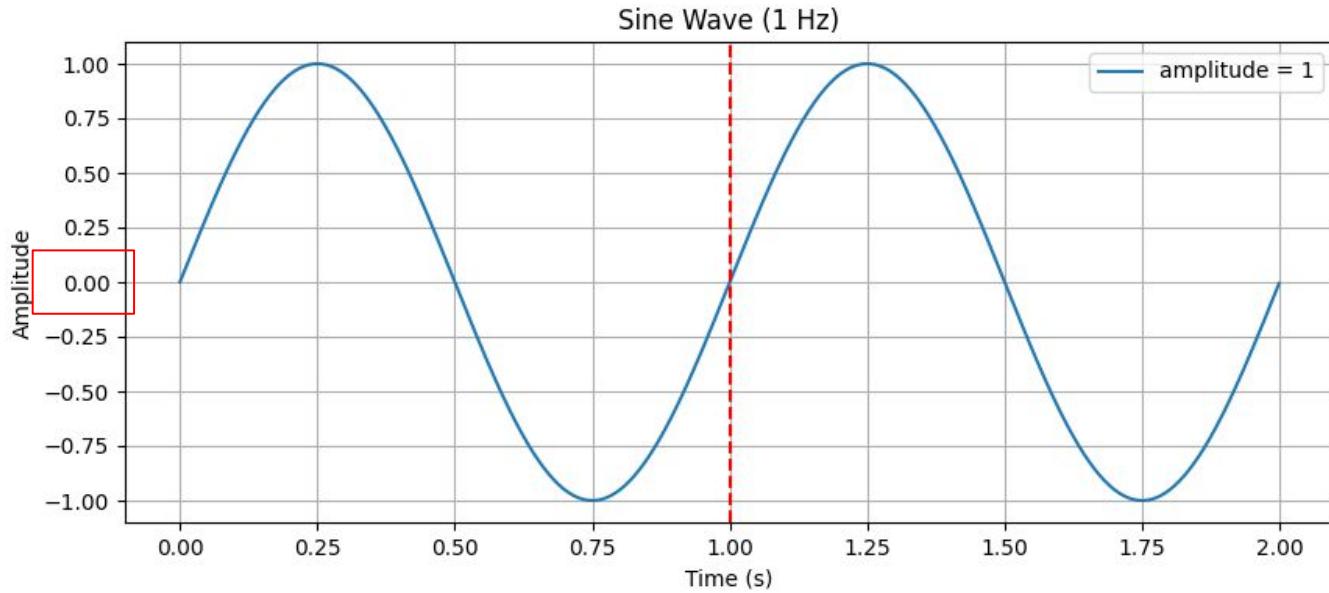
В чем мы измеряем амплитуду?



Properties of Sound Waves: Decibel

В чём мы измеряем амплитуду?

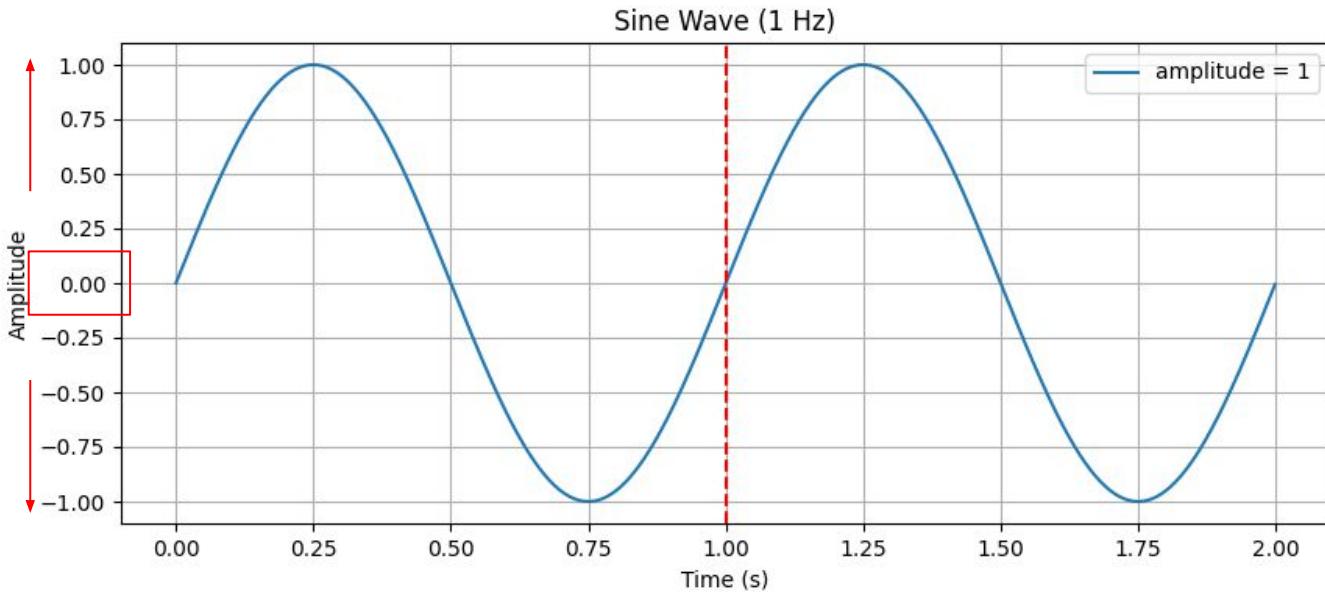
Отклонение от
атмосферного
давления



Properties of Sound Waves: Decibel

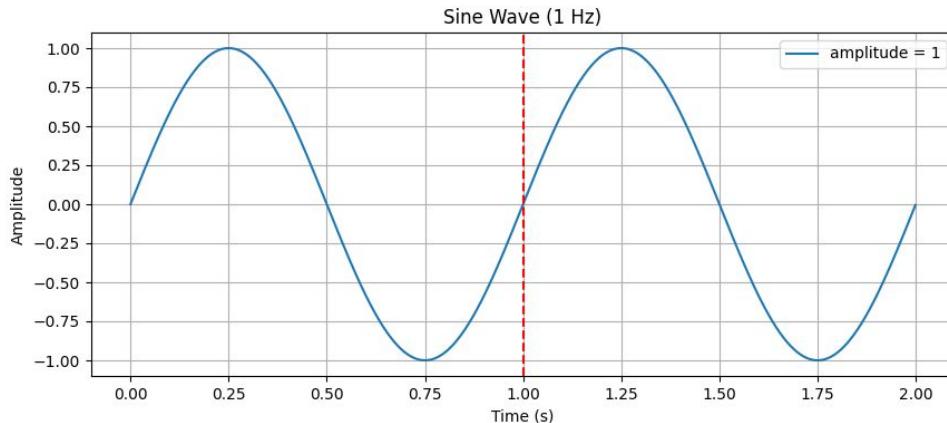
В чём мы измеряем амплитуду?

Отклонение от
атмосферного
давления



Properties of Sound Waves: Decibel

Амплитуда [Па?]



В чем мы измеряем амплитуду?

Атмосферное давление = 101 325 Па

Тогда и отклонение от атмосферного давления измеряется в Па?



Да, но это неудобно.

Properties of Sound Waves: Decibel

Почему неудобно использовать Па в качестве измерений?

1 причина. Диапазон воспринимаемых человеком звуковых давлений огромен.

- Порог слышимости: ~ 0.00002 Па (20 мкПа)

Пример: Тиканье часов или шепот.

- Разговорная речь: ~ 0.02 - 0.06 Па
- Болевой порог: ~ 20 Па

Пример: Сирена, громкий взрыв.

Ощущения: Вызывает физическую боль.

- Смертельный уровень: > 20 Па

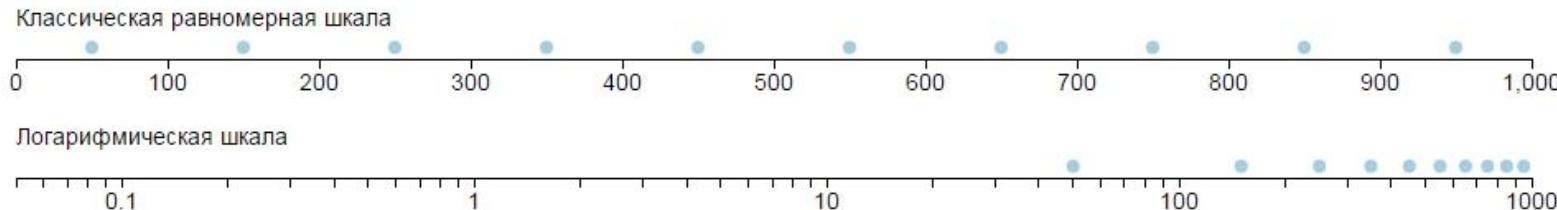
Последствия: Может вызвать потерю сознания, разрыв барабанных перепонок и даже смерть.

диапазон от 0.00002 Па до 20 Па! Это отношение 1 к 1,000,000! Оперировать такими числами очень неудобно.

Properties of Sound Waves: Decibel

Почему неудобно использовать Па в качестве измерений?

2 причина. Человеческое ухо воспринимает звук логарифмически.



Мы лучше замечаем изменения в тихих звуках, чем такие же абсолютные изменения в громких звуках.

Допустим, в тихой комнате зашумел компьютер, и уровень звука поднялся. Вы это сразу заметите.

Теперь на шумной улице проехал еще один автомобиль, и уровень звука тоже поднялся на примерно ту же величину. Вы это почти не заметите, хотя абсолютное изменение давления (в Паскалях) во втором случае в тысячи раз больше!

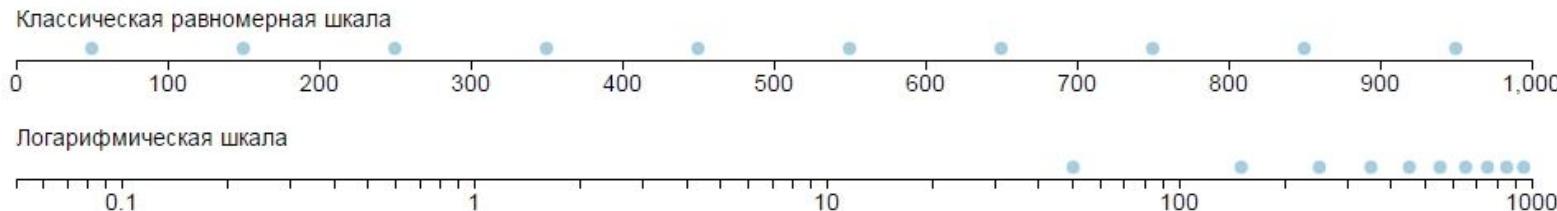
Почему?

Потому что для нашего уха важно не на сколько Паскалей, а во сколько раз изменилось давление.

Properties of Sound Waves: Decibel

Почему неудобно использовать Па в качестве измерений?

2 причина. Человеческое ухо воспринимает звук логарифмически.



Мы лучше замечаем изменения в тихих звуках, чем такие же абсолютные изменения в громких звуках.

Допустим, в тихой комнате (0.000632 Па) зашумел компьютер, и уровень звука поднялся до 0.00112 Па. Вы это **сразу заметите**, хотя разница равна **0.000487 Па**.

Теперь на шумной улице (0.2 Па) проехал еще один автомобиль, и уровень звука **тоже поднялся** до 0.35 Па. Вы это **почти не заметите**, хотя абсолютное изменение давления (в Паскалях) во втором случае в тысячи раз больше (**0.149 Па**)!

Почему?

Потому что для нашего уха важно не на сколько Паскалей, а **во сколько раз** изменилось давление.

Properties of Sound Waves: Decibel

Почему неудобно использовать Па в качестве измерений?

1 причина. Диапазон воспринимаемых человеком звуковых давлений огромен.

2 причина. Человеческое ухо воспринимает звук логарифмически.



Давайте введем логарифмическую относительную величину → Децибел (Decibel).

Что такое Децибел ?

Давайте возьмем **самый тихий звук**, который только может воспринимать человек за значение, от которого мы будем отталкиваться. Это 0.00002 Па.

$$L_p = 20 \cdot \log_{10} \frac{P}{P_0}$$

считаем
относительное
изменение

$$P_0 = 0.00002 \text{ Па}$$

Properties of Sound Waves: Decibel

Почему неудобно использовать Па в качестве измерений?

1 причина. Диапазон воспринимаемых человеком звуковых давлений огромен.

2 причина. Человеческое ухо воспринимает звук логарифмически.



Давайте введем логарифмическую относительную величину → Децибел (Decibel).

Что такое Децибел ?

Давайте возьмем **самый тихий звук**, который только может воспринимать человек за значение, от которого мы будем отталкиваться. Это 0.00002 Па.

$$L_p = 20 \cdot \log_{10} \frac{P}{P_0}$$

переводим в лог
шкалу для
удобства

$$P_0 = 0.00002 \text{ Pa}$$

Properties of Sound Waves: Decibel

Почему неудобно использовать Па в качестве измерений?

1 причина. Диапазон воспринимаемых человеком звуковых давлений огромен.

2 причина. Человеческое ухо воспринимает звук логарифмически.



Давайте введем логарифмическую относительную величину → Децибел (Decibel).

Что такое Децибел ?

Давайте возьмем **самый тихий звук**, который только может воспринимать человек за значение, от которого мы будем отталкиваться. Это 0.00002 Па.

$$L_p = \boxed{20} \cdot \log_{10} \frac{P}{P_0}$$

$$P_0 = 0.00002 \text{ Pa}$$

Почему тут 20 стоит?

Чтобы разобраться с этим вопросом, давайте разберемся с понятиями мощности, энергии и интенсивности.

Properties of Sound Waves: Power, Energy and Intensity

Изначально децибелы были введены для измерения мощности, так как исторически в телефонии и радиотехнике инженеров интересовала **мощность** (P) — полная энергия в секунду, которую источник передает по линии.

Мощность измеряется в Ваттах (Вт).

Изначально была введена единица **Бел (Bel)** в честь Александра Грейама Белла.

$$L_p[\text{Bel}] = \log_{10} \frac{p_1}{p_2}$$

(Здесь p - мощность)

Она определялась для измерения **уменьшения мощности** сигнала в телефонии (затухания в длинных линиях связи).

Если мощность сигнала уменьшилась в 100 раз, то затухание составляло $\log_{10}(100) = 2$ B (Бела).

На практике инженерам нужно было измерять **гораздо более мелкие изменения** мощности.

$$1 \text{ dB} = 0.1 \cdot \text{Bel}$$

Поэтому люди начали использовать **децибел** (дБ).

$$L_p[\text{dB}] = \boxed{10} \cdot \log_{10} \frac{p_1}{p_2}$$

Это то откуда в формуле появился коэффициент 10.

Properties of Sound Waves: Power, Energy and Intensity

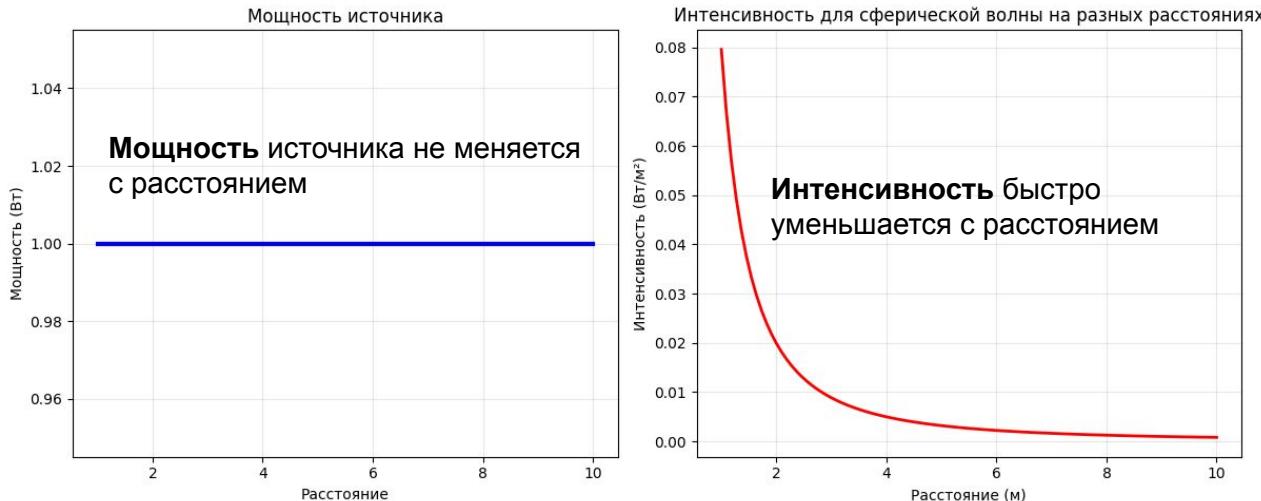
Но когда мы переходим к акустике и распространению звука в пространстве, возникает важный вопрос: **"Где именно?"**

Мощность источника — это его общая "сила", но то, что действительно достигает уха или микрофона, — это **интенсивность** (I) — мощность, проходящая через единицу площади.

Интенсивность измеряется в Ваттах на квадратный метр ($\text{Вт}/\text{м}^2$).

$$I = \frac{P}{S}$$

- I — интенсивность ($\text{Вт}/\text{м}^2$)
- P — мощность источника (Вт)
- S — площадь (м^2)



Поэтому начали измерять интенсивность и тоже в децибелах.

$$L_I = 10 \cdot \log_{10} \frac{I}{I_0}$$

Properties of Sound Waves: Power, Energy and Intensity

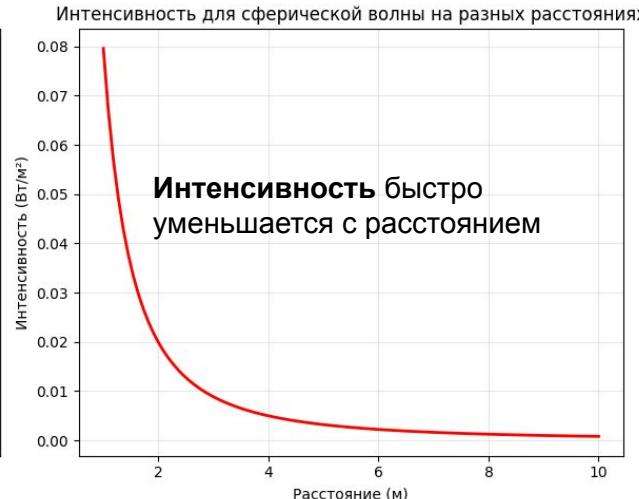
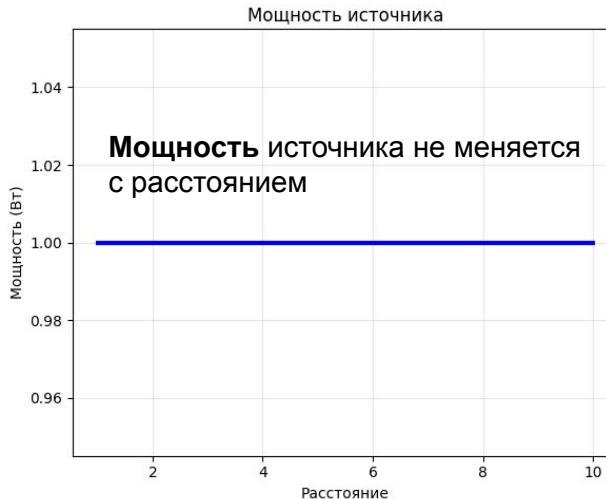
Но когда мы переходим к акустике и распространению звука в пространстве, возникает важный вопрос: **"Где именно?"**

Мощность источника — это его общая "сила", но то, что действительно достигает уха или микрофона, — это **интенсивность** (I) — мощность, проходящая через единицу площади.

Интенсивность измеряется в Ваттах на квадратный метр ($\text{Вт}/\text{м}^2$).

$$I = \frac{P}{S}$$

- I — интенсивность ($\text{Вт}/\text{м}^2$)
- P — мощность источника (Вт)
- S — площадь (м^2)



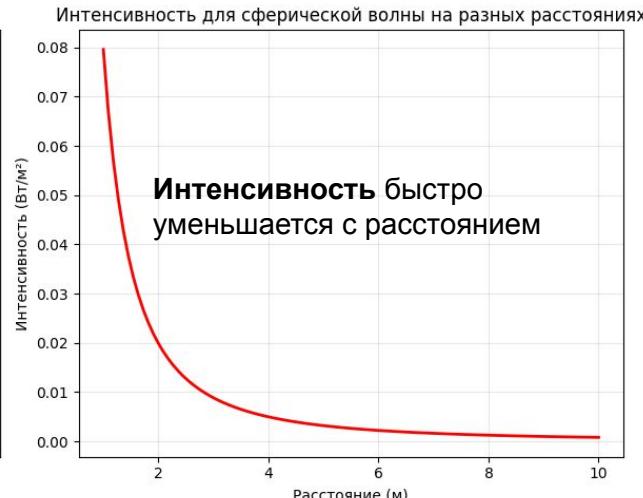
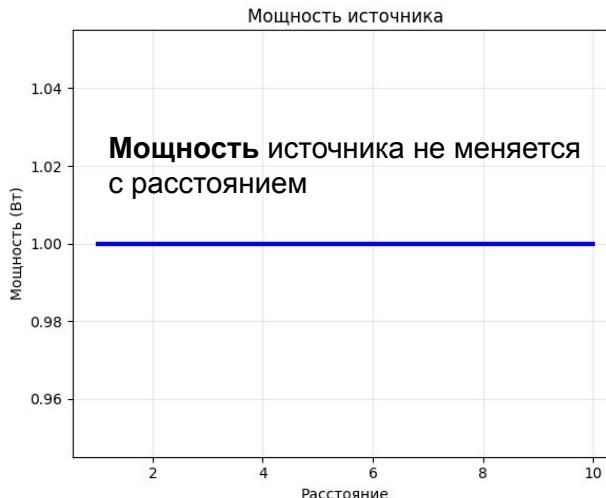
И от интенсивности к давлению мы перешли, потому что давление — это то, что непосредственно измеряют микрофоны и воспринимают уши!

Properties of Sound Waves: Power, Energy and Intensity

Но когда мы переходим к акустике и распространению звука в пространстве, возникает важный вопрос: **"Где именно?"**

Мощность источника — это его общая "сила", но то, что действительно достигает уха или микрофона, — это **интенсивность (I)** — мощность, проходящая через единицу площади.

Интенсивность измеряется в Ваттах на квадратный метр ($\text{Вт}/\text{м}^2$).



Разбираемся.

Еще раз, у нас есть две сущности: **источник** излучающий волны и **сами волны**.

Мощность определяет полную энергию в секунду, которую источник излучает во все направления.

Давление возникает в поле на определенном расстоянии от источника.

Интенсивность определяет изменение энергии от расстояния, и прямо пропорциональна амплитуде давления в точке, через которую распространяется энергия источника, которое порождает звуковые волны.

Properties of Sound Waves: Decibel

Почему неудобно использовать Па в качестве измерений?

1 причина. Диапазон воспринимаемых человеком звуковых давлений огромен.

2 причина. Человеческое ухо воспринимает звук логарифмически.



Давайте введем логарифмическую относительную величину → Децибел (Decibel).

Что такое Децибел ?

Давайте возьмем **самый тихий звук**, который только может воспринимать человек за значение, от которого мы будем отталкиваться. Это 0.00002 Па.

$$L_p = \boxed{20} \cdot \log_{10} \frac{P}{P_0}$$

$$P_0 = 0.00002 \text{ Pa}$$

Почему тут 20 стоит?

$$\boxed{L_I = 10 \cdot \log_{10} \frac{I}{I_0}}$$
$$I \sim P^2$$

$$10 \cdot \log_{10} \frac{I}{I_0} = 10 \cdot \log_{10} \frac{P^2}{P_0^2} = 10 \cdot \log_{10} \left(\frac{P}{P_0} \right)^2 = 10 \cdot 2 \log_{10} \frac{P}{P_0} = 20 \cdot \log_{10} \frac{P}{P_0}$$

Properties of Sound Waves: Decibel

Почему неудобно использовать Па в качестве измерений?

1 причина. Диапазон воспринимаемых человеком звуковых давлений огромен.

2 причина. Человеческое ухо воспринимает звук логарифмически.



Давайте введем логарифмическую относительную величину → Децибел (Decibel).

Что такое Децибел ?

Давайте возьмем **самый тихий звук**, который только может воспринимать человек за значение, от которого мы будем отталкиваться. Это 0.00002 Па.

$$L_p = \boxed{20} \cdot \log_{10} \frac{P}{P_0}$$

$$P_0 = 0.00002 \text{ Pa}$$

Почему тут 20 стоит?

$$L_I = 10 \cdot \log_{10} \frac{I}{I_0}$$

$$I \sim P^2$$

$$10 \cdot \log_{10} \frac{I}{I_0} = 10 \cdot \log_{10} \frac{P^2}{P_0^2} = 10 \cdot \log_{10} \left(\frac{P}{P_0} \right)^2 = 10 \cdot 2 \log_{10} \frac{P}{P_0} = 20 \cdot \log_{10} \frac{P}{P_0}$$

Properties of Sound Waves: Decibel

Описание звука	Давление (Па)	Уровень давления (SPL - Sound Pressure Level) (дБ)
Порог слышимости	0.00002	0 дБ
Тихая комната	0.0002	20 дБ
Шепот	0.002	40 дБ
Спокойный разговор	0.02	60 дБ
Шумная улица	0.2	80 дБ
Отбойный молоток	2	100 дБ
Болевой порог	20	120 дБ
Реактивный двигатель	200	140 дБ

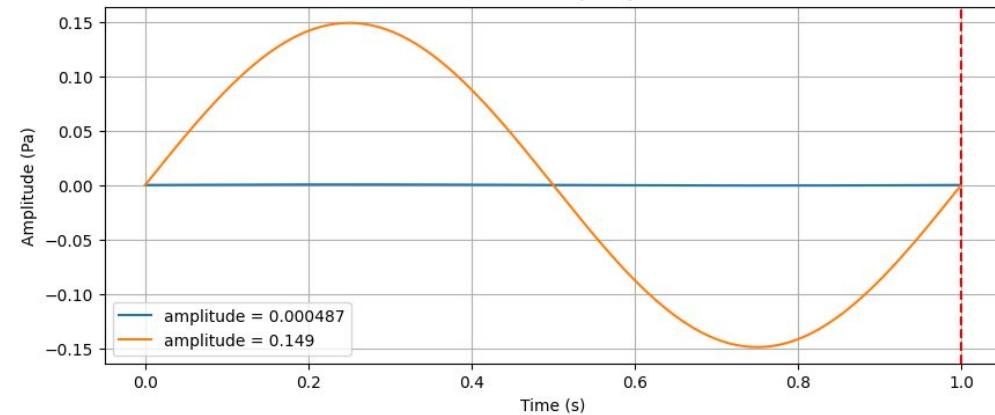
Шкала децибел для реального звука начинается с 0 дБ, что соответствует самому тихому звуку, который может услышать человек.

$$L_p = 20 \cdot \log_{10} \frac{P}{P_0}$$

$$P_0 = 0.00002 \text{ Па}$$

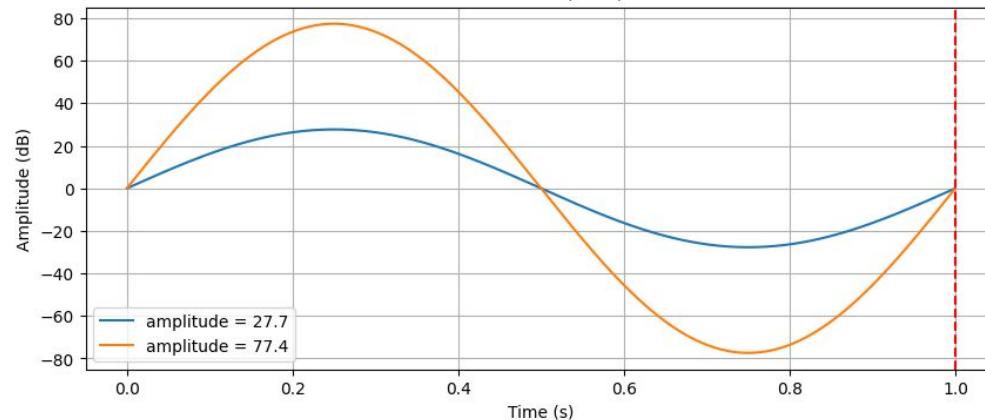
Properties of Sound Waves: Decibel

Sine Wave (1 Hz)



в Паскалях

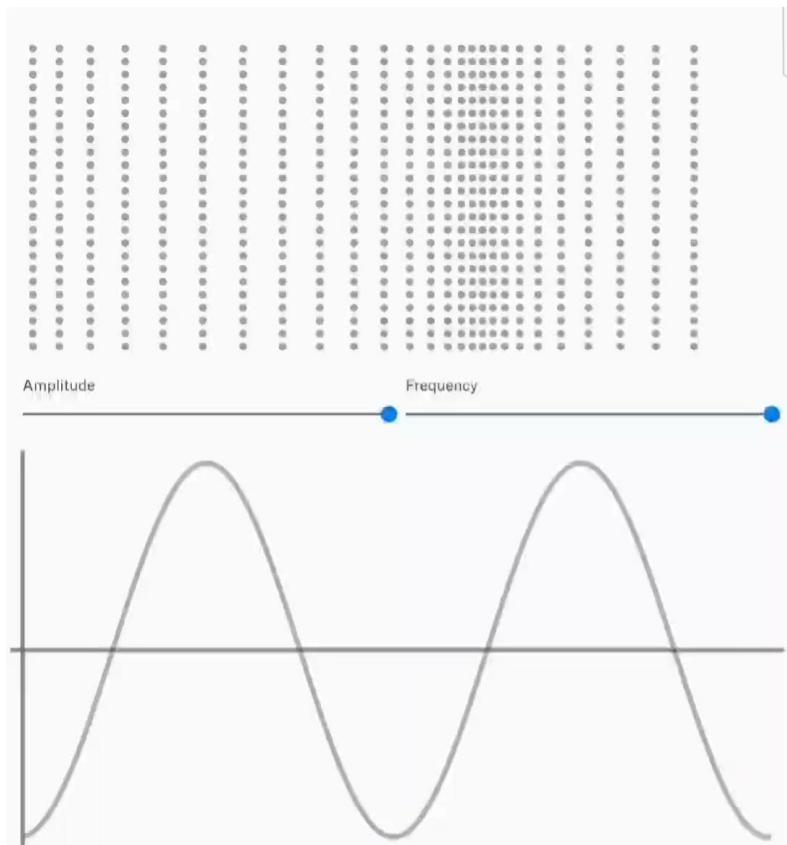
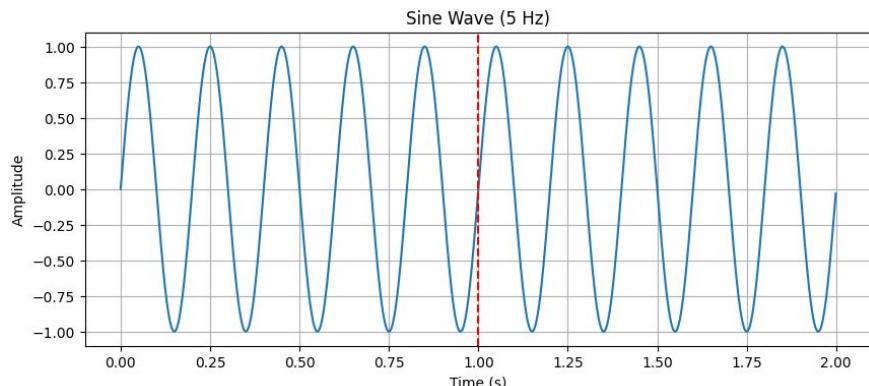
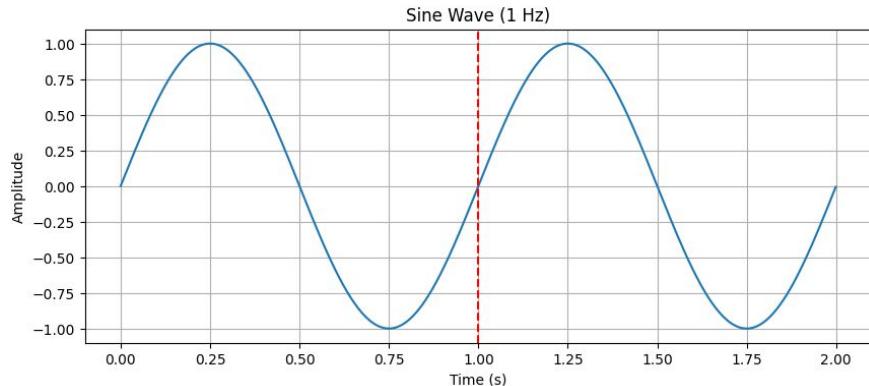
Sine Wave (1 Hz)



в Децибелах

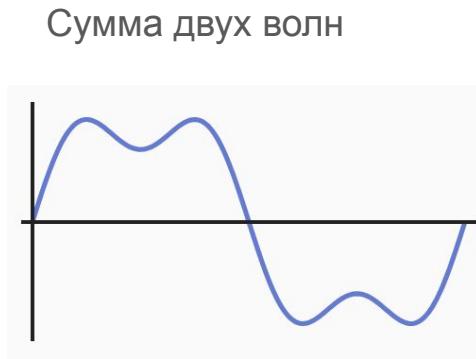
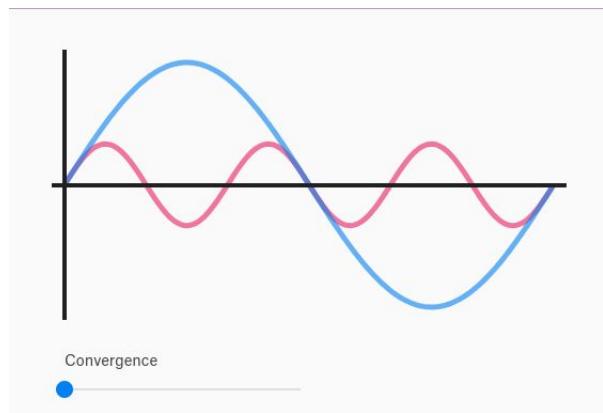
Properties of Sound Waves: Frequency

Frequency [Hz]

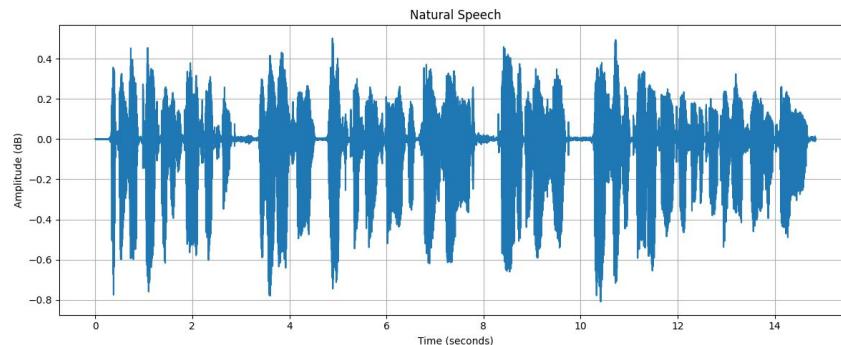


<https://pudding.cool/2018/02/waveforms/>

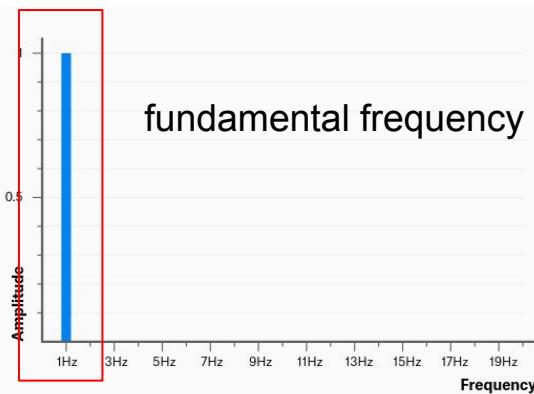
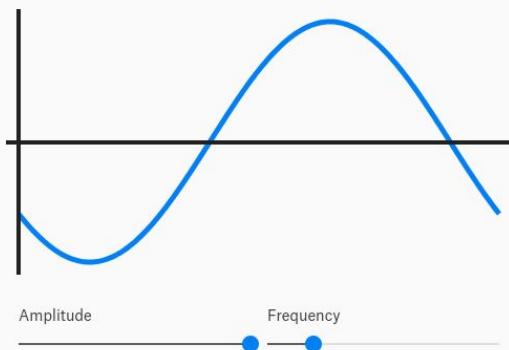
Properties of Sound Waves: Waves summation



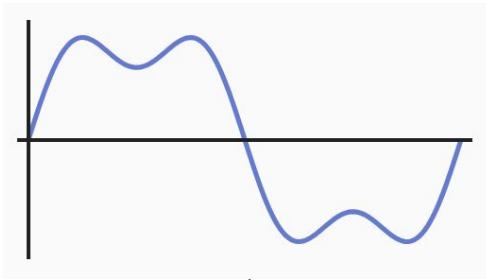
Как выглядит
настоящий звук,
например, тут
человеческая речь



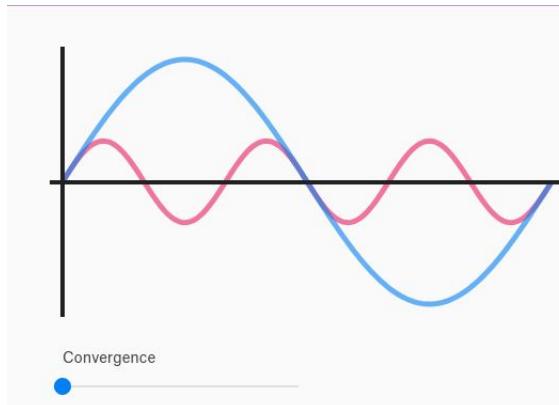
Properties of Sound Waves: Harmonics



Properties of Sound Waves: Harmonics



Сумма двух волн



Гармоники — это дополнительные частоты, создаваемые определёнными формами волн. Гармоники всегда кратны основной частоте.

Обозначения:

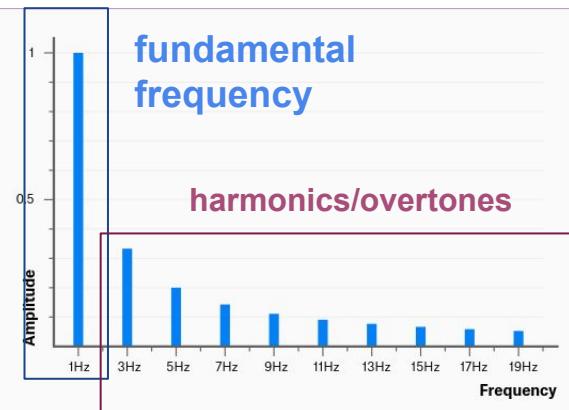
1f - Fundamental frequency, 1st Harmonic

2f - 1st Overtone, 2nd Harmonic...etc...

3f - 2st Overtone, 3rd Harmonic...etc...

Это основной тон, который вы слышите.

Если сыграть одну и ту же ноту (одинаковую основную частоту)



почему они будут звучать совершенно по-разному?

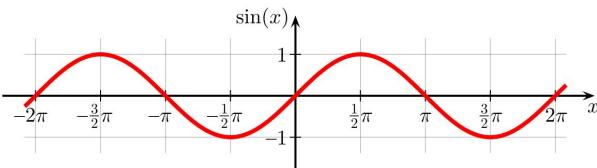
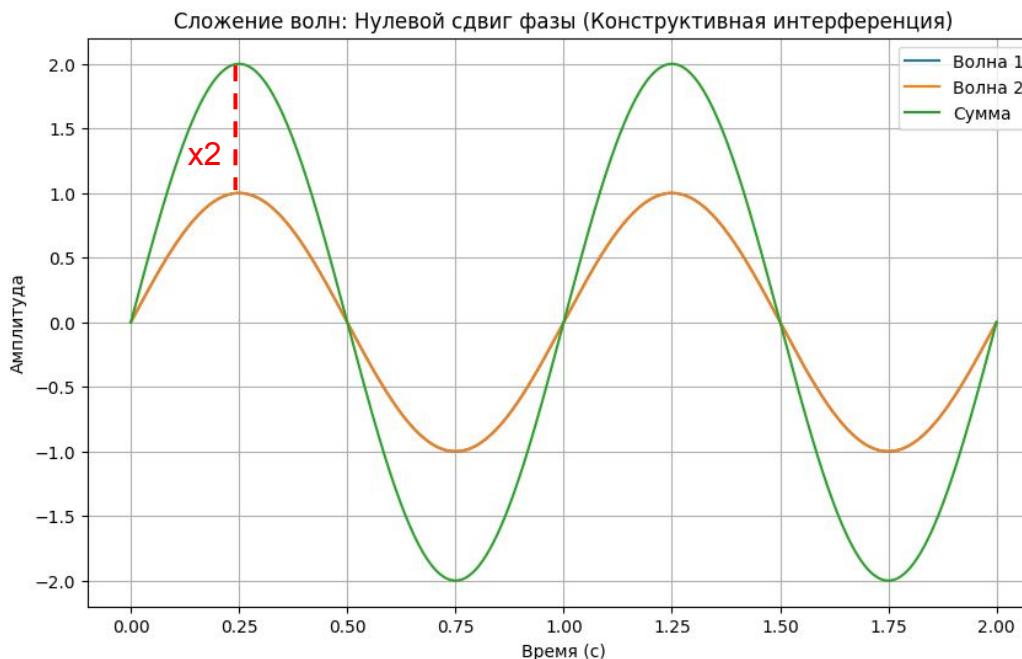
Уникальное сочетание и относительная громкость гармоник дают уникальный **тембр** звуку.

Properties of Sound Waves: Phase.

$$y(t) = A \cdot \sin(2\pi ft + \phi)$$

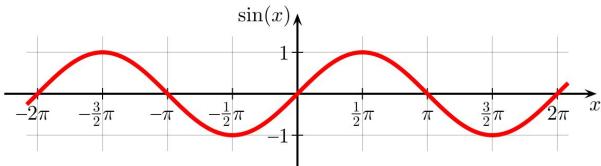
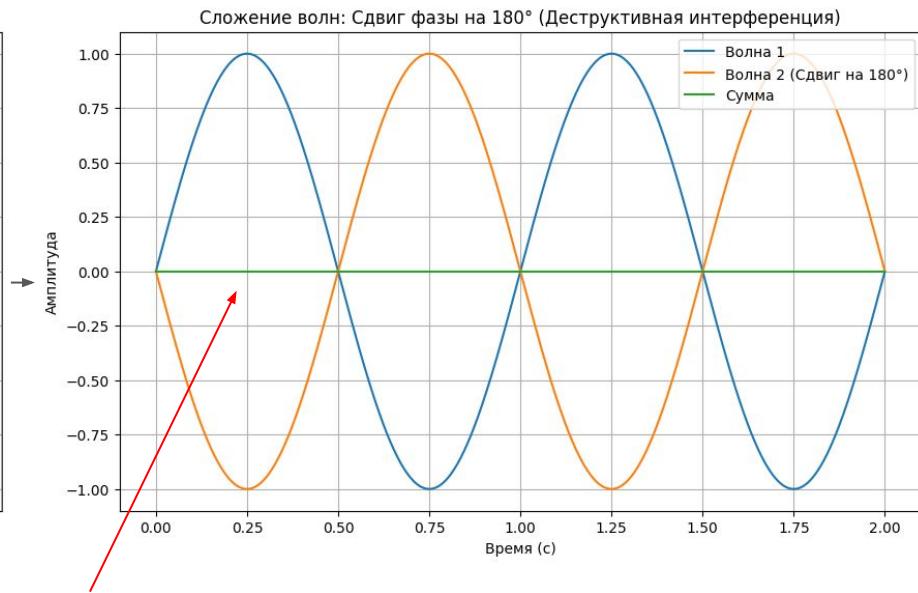
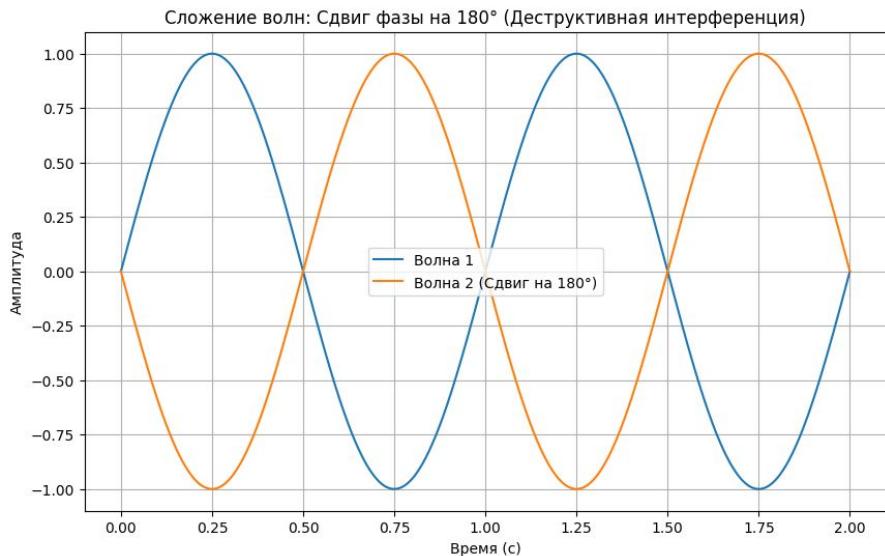
$$\phi = 0$$

Амплитуда
увеличилась в
два раза



Properties of Sound Waves: Phase. Noise Canceling.

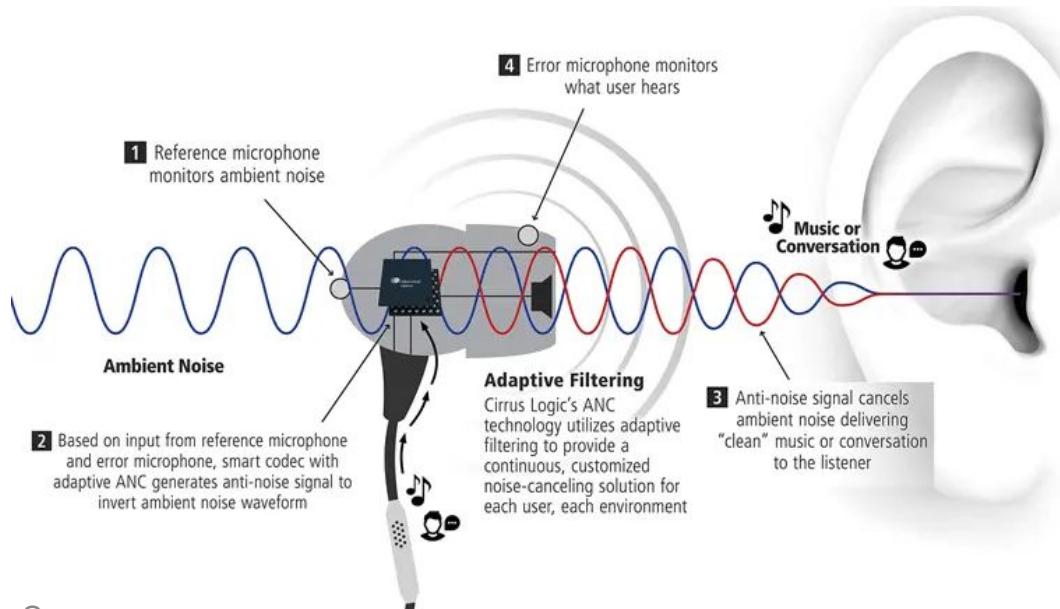
$$\phi = 180^\circ$$



Они взаимно уничтожат друг друга!

Это буквально как работает
шумоподавление в наушниках.

Properties of Sound Waves: Phase. Noise Canceling.



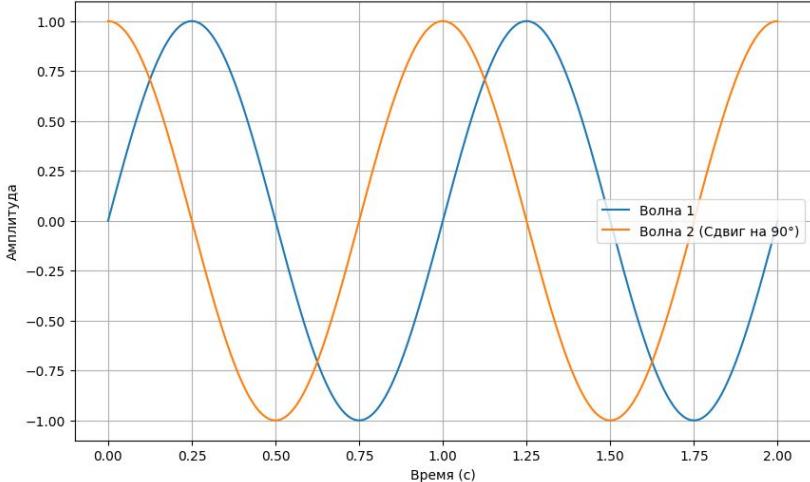
Source:

<https://ash-asia.zendesk.com/hc/en-us/articles/41558869165721-How-do-Active-Noise-Canceling-A-N-C-headphones-work>

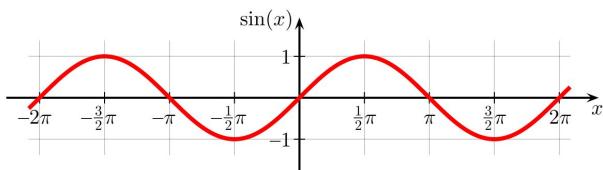
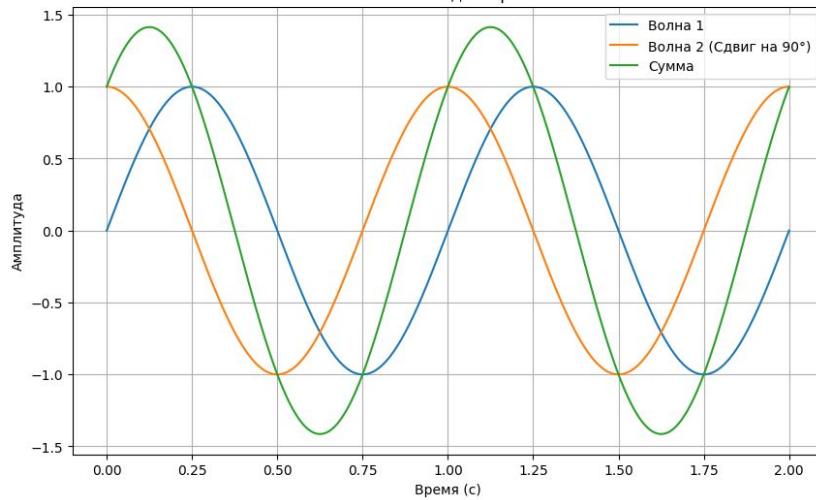
Properties of Sound Waves: Phase.

$$\phi = 90^\circ$$

Сложение волн: Сдвиг фазы на 90°



Сложение волн: Сдвиг фазы на 90°

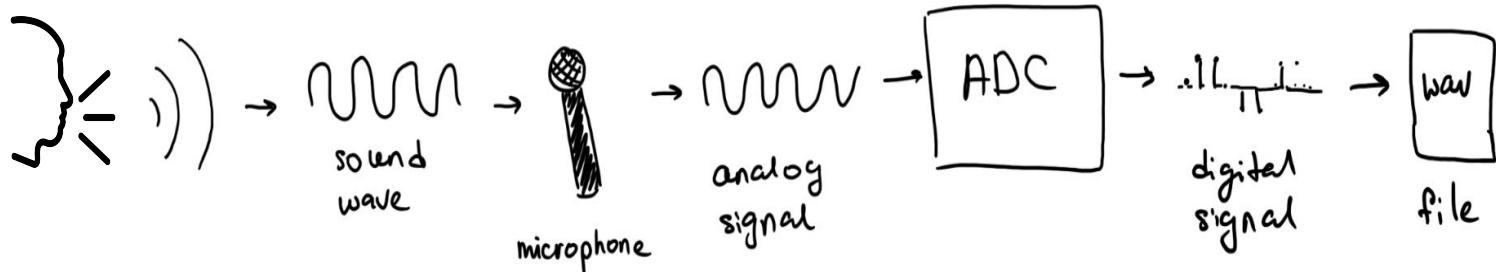


Discretization of an audio signal

Lecture: Intro to Speech Processing



Discretization Pipeline: Analogue to Digital Converter



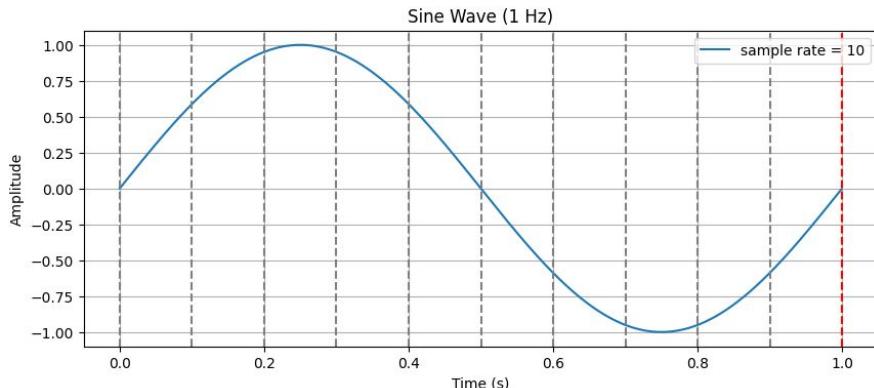
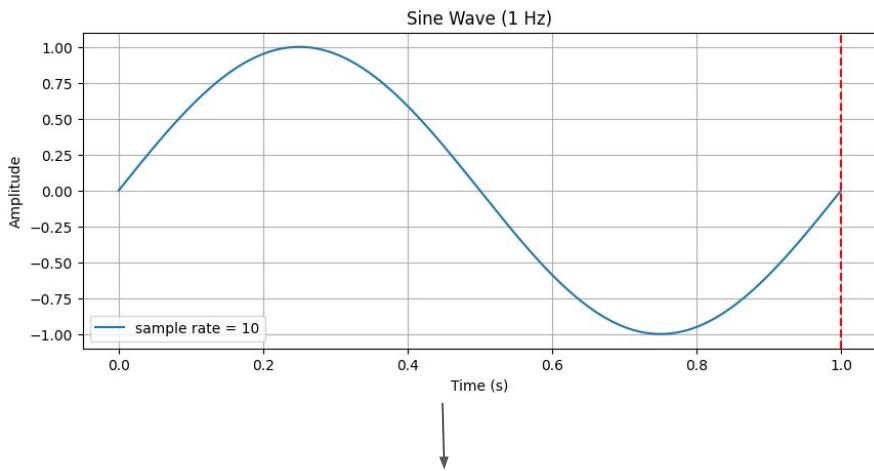
Принцип работы микрофона:

1. Диафрагма (очень тонкая, гибкая мембрана), которая находится внутри микрофона, подвергается воздействию звуковых колебаний воздуха. Когда звуковые волны достигают ее, она начинает колебаться из-за изменения давления воздуха.
2. Физическое движение диафрагмы преобразуется в электрический сигнал. Это основа работы микрофона, и именно этим отличаются разные типы микрофонов, т.е. тем как движение диафрагмы конвертируется в электрический сигнал.
3. Слабый электрический сигнал передается из микрофона по кабелю (обычно XLR или разъему) на усилитель или аудиоинтерфейс.



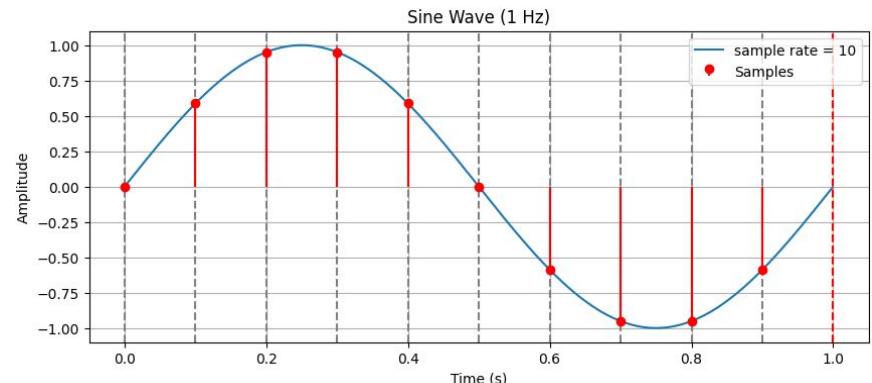
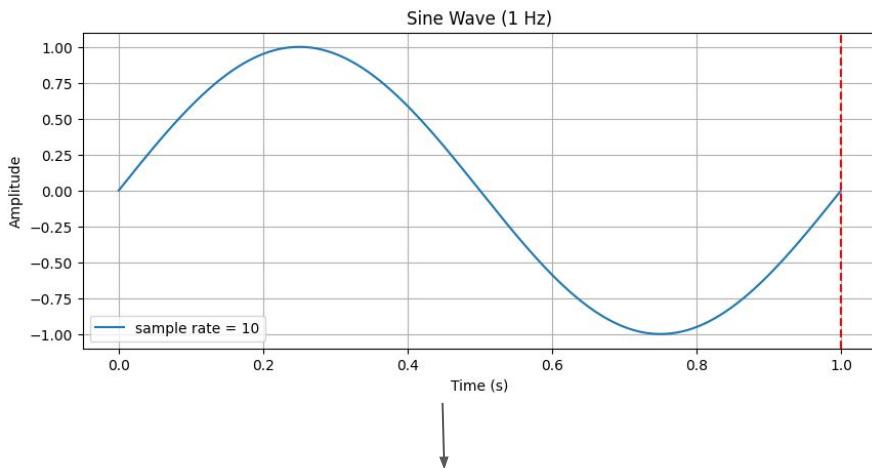
[Link to the source](#)

Analog to Digital Converter: Time Discretization



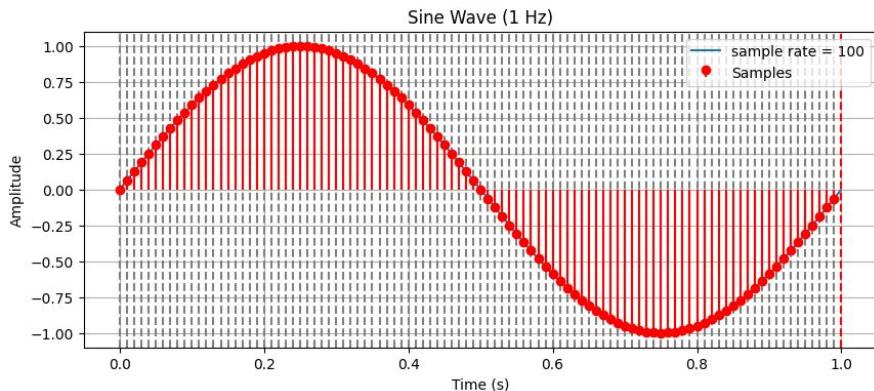
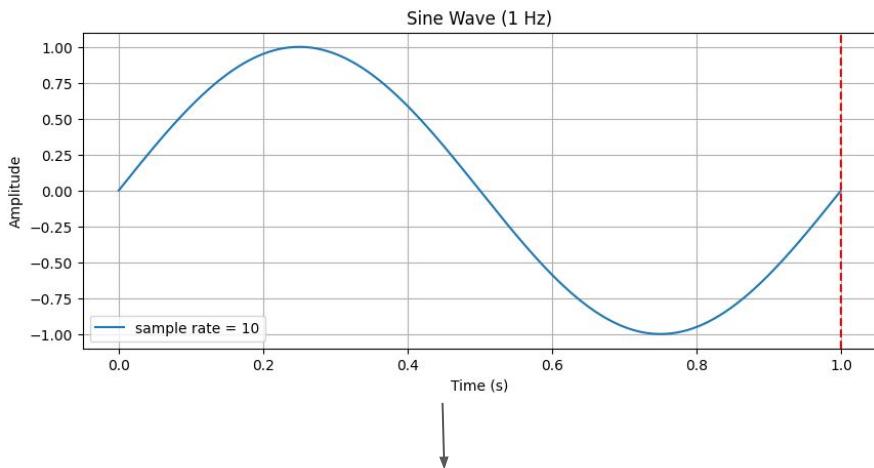
The **sampling rate** (also called sampling frequency) is the number of samples taken in one second and is measured in hertz (Hz).

Analog to Digital Converter: Time Discretization



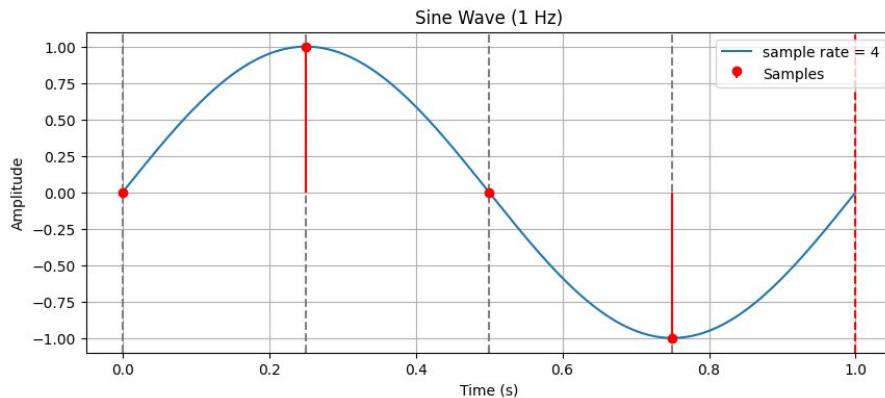
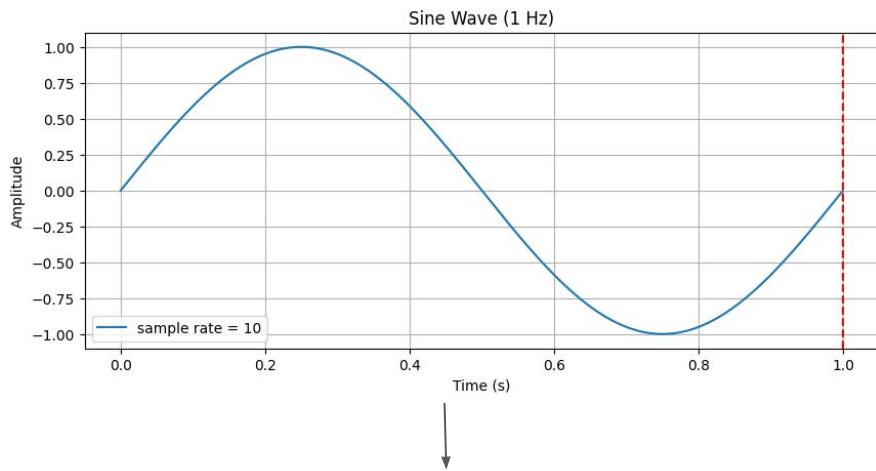
The **sampling rate** (also called sampling frequency) is the number of samples taken in one second and is measured in hertz (Hz).

Analog to Digital Converter: Time Discretization



The **sampling rate** (also called sampling frequency) is the number of samples taken in one second and is measured in hertz (Hz).

Analog to Digital Converter: Time Discretization



The **sampling rate** (also called sampling frequency) is the number of samples taken in one second and is measured in hertz (Hz).

How often should we sample to perfectly reconstruct a signal?

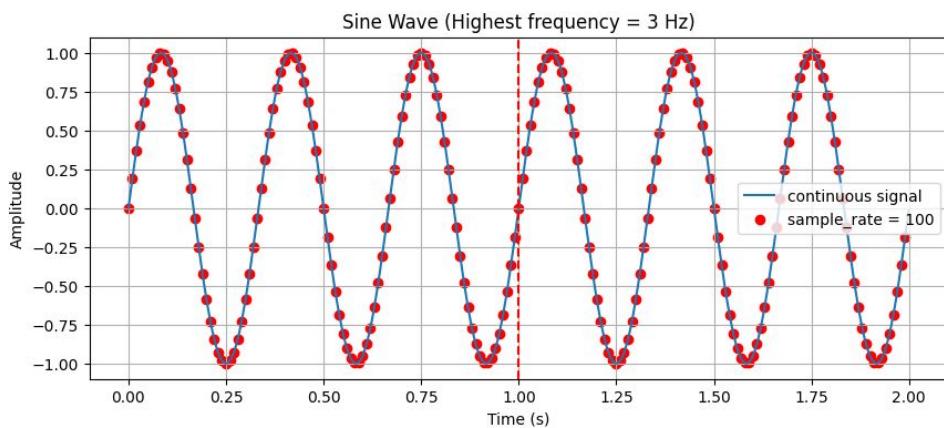
Shannon Nyquist Sampling Theorem

How often should we sample to perfectly reconstruct a signal?

Key Principle: To perfectly reconstruct a signal from its samples, you must sample at least **twice the highest frequency** present in the signal.

$$f_s \geq 2 \times f_{\text{ax}}$$

- f_s = sampling frequency
- f_{ax} = highest frequency in the signal



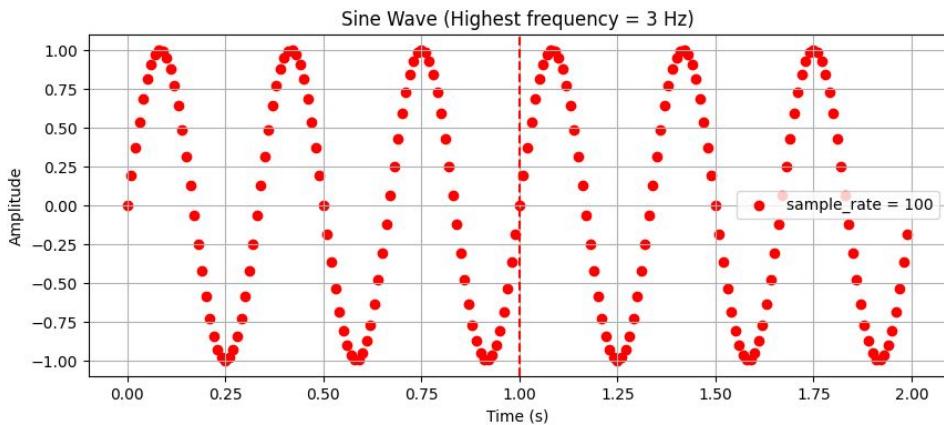
Shannon Nyquist Sampling Theorem

How often should we sample to perfectly reconstruct a signal?

Key Principle: To perfectly reconstruct a signal from its samples, you must sample at least **twice the highest frequency** present in the signal.

$$f_s \geq 2 \times f_{\text{ax}}$$

- f_s = sampling frequency
- f_{ax} = highest frequency in the signal



Можете ли вы
определить какой был
оригинальный сигнал
после дискретизации?

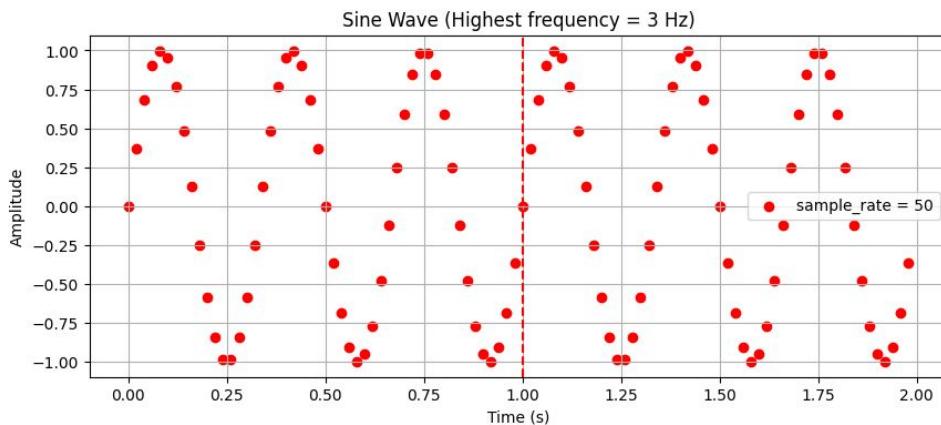
Shannon Nyquist Sampling Theorem

How often should we sample to perfectly reconstruct a signal?

Key Principle: To perfectly reconstruct a signal from its samples, you must sample at least **twice the highest frequency** present in the signal.

$$f_s \geq 2 \times f_{\text{ax}}$$

- f_s = sampling frequency
- f_{ax} = highest frequency in the signal



Можете ли вы
определить какой был
оригинальный сигнал
после дискретизации?

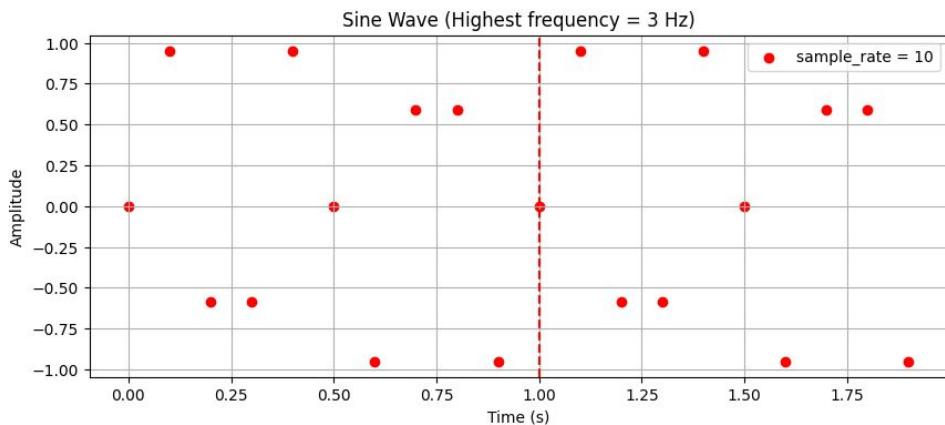
Shannon Nyquist Sampling Theorem

How often should we sample to perfectly reconstruct a signal?

Key Principle: To perfectly reconstruct a signal from its samples, you must sample at least **twice the highest frequency** present in the signal.

$$f_s \geq 2 \times f_{\text{ax}}$$

- f_s = sampling frequency
- f_{ax} = highest frequency in the signal



Можете ли вы
определить какой был
оригинальный сигнал
после дискретизации?

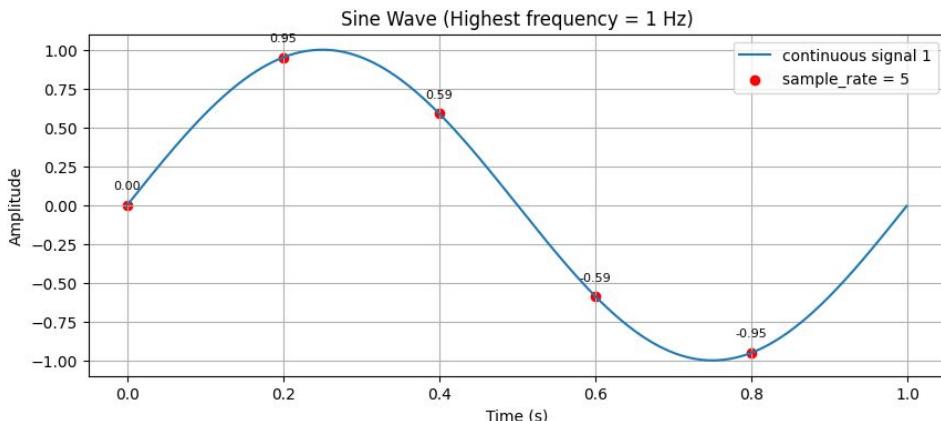
Shannon Nyquist Sampling Theorem

How often should we sample to perfectly reconstruct a signal?

Key Principle: To perfectly reconstruct a signal from its samples, you must sample at least **twice the highest frequency** present in the signal.

$$f_s \geq 2 \times f_{\text{ax}}$$

- f_s = sampling frequency
- f_{ax} = highest frequency in the signal



Или скорее ответим на вопрос:
Есть ли еще какие то синусоиды, которые будут также проходить по этим точкам?

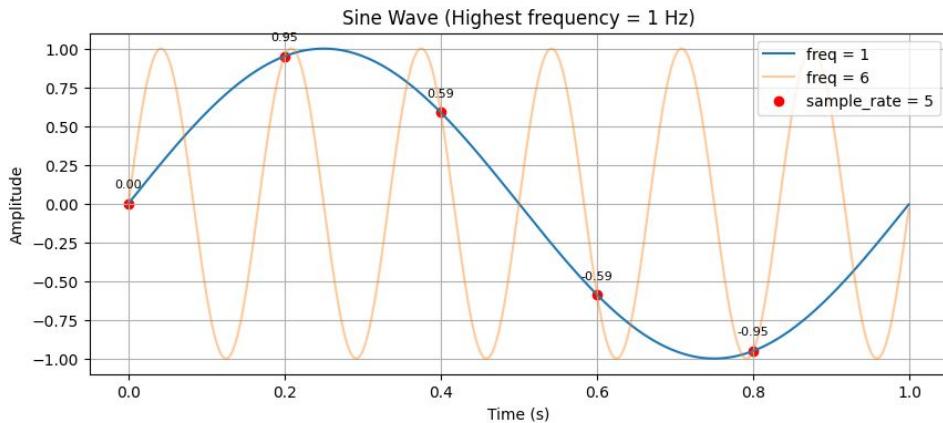
Shannon Nyquist Sampling Theorem

How often should we sample to perfectly reconstruct a signal?

Key Principle: To perfectly reconstruct a signal from its samples, you must sample at least **twice the highest frequency** present in the signal.

$$f_s \geq 2 \times f_{\text{ax}}$$

- f_s = sampling frequency
- f_{ax} = highest frequency in the signal



Или скорее ответим на вопрос:
Есть ли еще какие то синусоиды, которые будут также проходить по этим точкам?

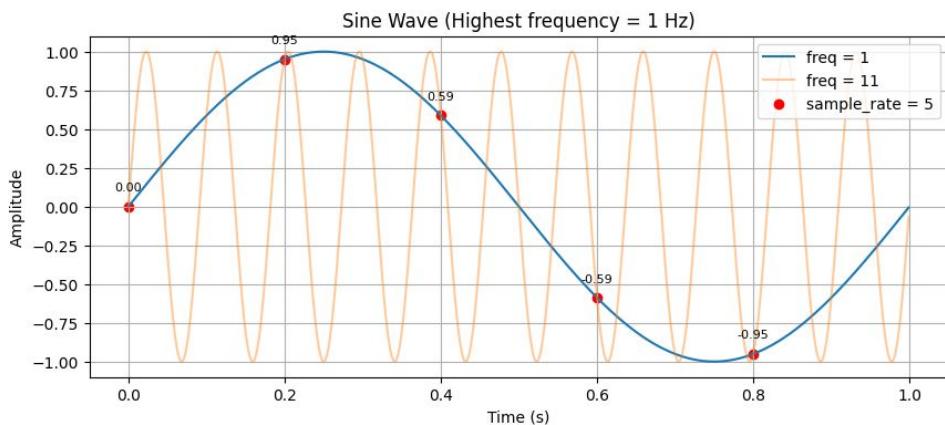
Shannon Nyquist Sampling Theorem

How often should we sample to perfectly reconstruct a signal?

Key Principle: To perfectly reconstruct a signal from its samples, you must sample at least **twice the highest frequency** present in the signal.

$$f_s \geq 2 \times f_{\text{ax}}$$

- f_s = sampling frequency
- f_{ax} = highest frequency in the signal



Или скорее ответим на вопрос:
Есть ли еще какие то синусоиды, которые будут также проходить по этим точкам?

Предположительно,
правильные ответы:

$$y = A \cdot \sin(2\pi \cdot 1 \cdot t)$$

$$y = A \cdot \sin(2\pi \cdot 6 \cdot t)$$

$$y = A \cdot \sin(2\pi \cdot 11 \cdot t)$$

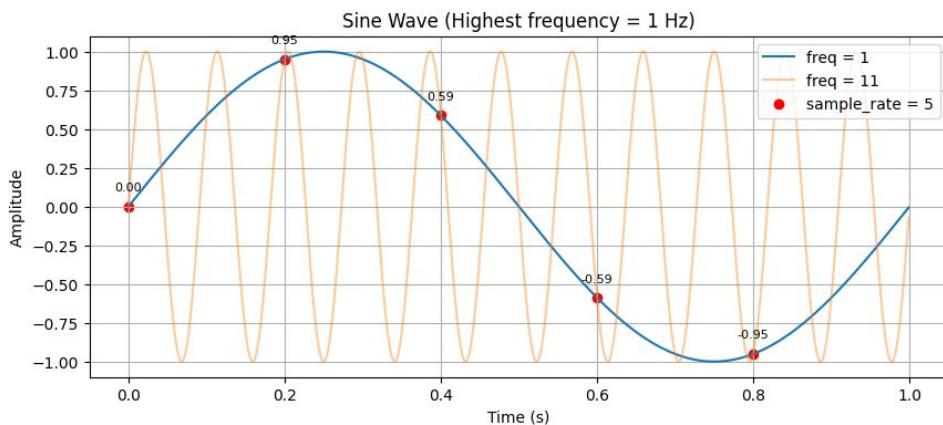
Shannon Nyquist Sampling Theorem

How often should we sample to perfectly reconstruct a signal?

Key Principle: To perfectly reconstruct a signal from its samples, you must sample at least **twice the highest frequency** present in the signal.

$$f_s \geq 2 \times f_{\text{ax}}$$

- f_s = sampling frequency
- f_{ax} = highest frequency in the signal



Или скорее ответим на вопрос:
Есть ли еще какие то синусоиды, которые будут также проходить по этим точкам?

Предположительно,
правильные ответы:

$$y = A \cdot \sin(2\pi \cdot 1 \cdot t)$$

$$y = A \cdot \sin(2\pi \cdot 6 \cdot t)$$

$$y = A \cdot \sin(2\pi \cdot 11 \cdot t)$$

Our original signal frequency is the lowest frequency we can guess.

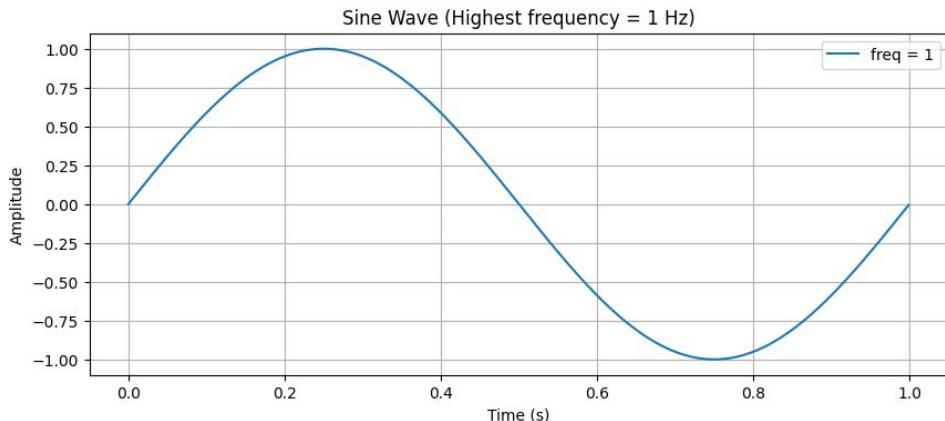
Shannon Nyquist Sampling Theorem

How often should we sample to perfectly reconstruct a signal?

Key Principle: To perfectly reconstruct a signal from its samples, you must sample at least **twice the highest frequency** present in the signal.

$$f_s \geq 2 \times f_{\text{max}}$$

- f_s = sampling frequency
- f_{max} = highest frequency in the signal



Что будет, если мы будем сэмплировать частотой меньше 2 Hz?

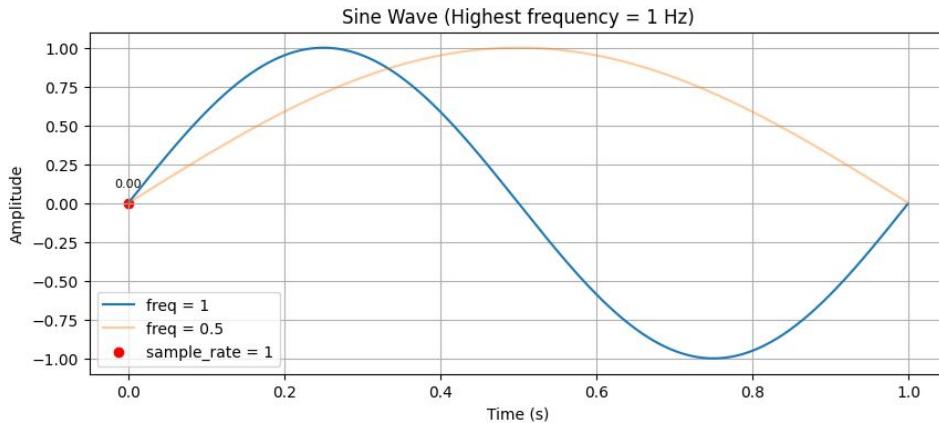
Shannon Nyquist Sampling Theorem

How often should we sample to perfectly reconstruct a signal?

Key Principle: To perfectly reconstruct a signal from its samples, you must sample at least **twice the highest frequency** present in the signal.

$$f_s \geq 2 \times f_{\text{ax}}$$

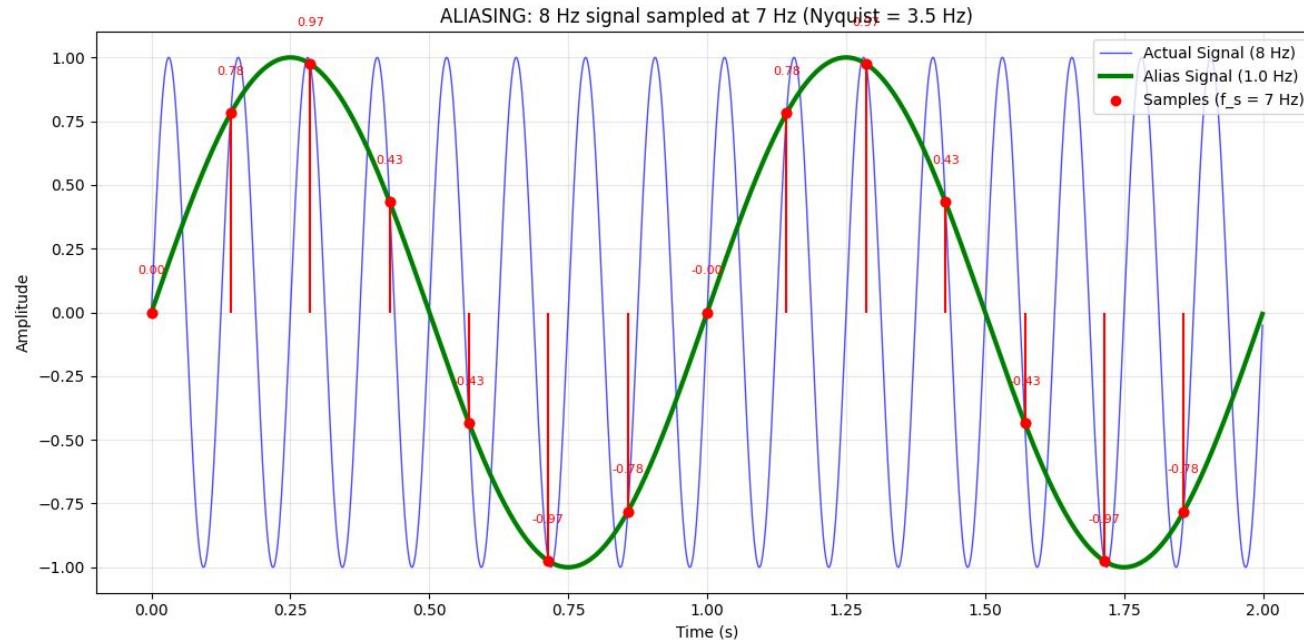
- f_s = sampling frequency
- f_{ax} = highest frequency in the signal



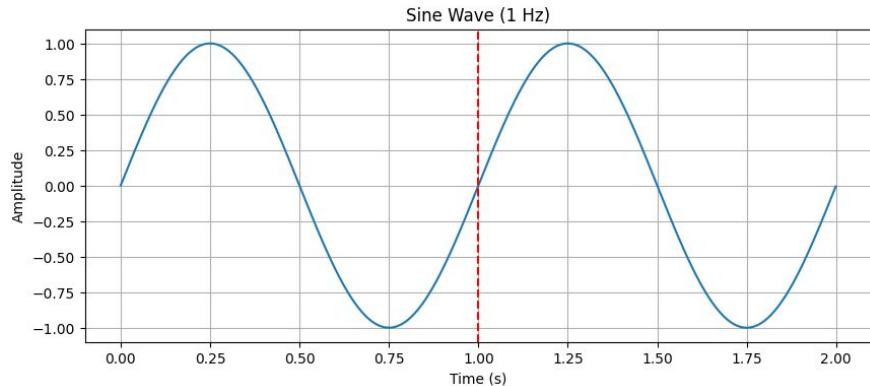
Найдется такая частота сигнала, которая будет меньше частоты оригинального сигнала.

Aliasing

Aliasing это тип искажения или ошибки, возникающий, когда сигнал дискретизируется с частотой, которая слишком низкая для захвата его самых высоких частот.

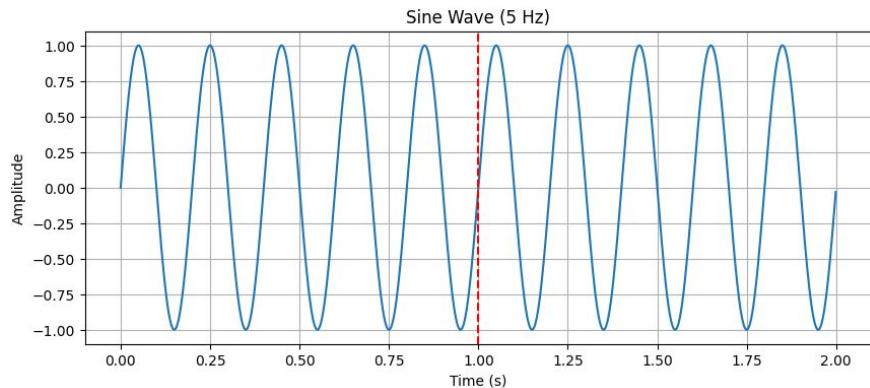


Shannon Nyquist Sampling Theorem



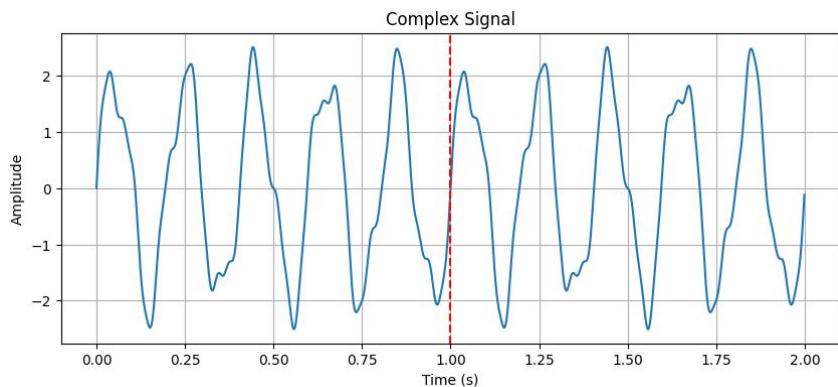
1 step. Find the highest frequency present in the signal

1 Hz - the highest frequency



5 Hz - the highest frequency

Shannon Nyquist Sampling Theorem



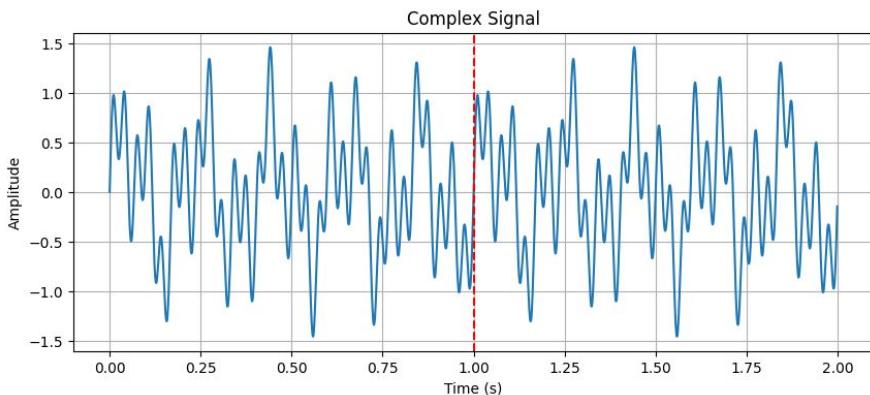
$$x(t) = 2 \cdot \sin(2\pi \cdot 5t) + 0.5 \cdot \sin(2\pi \cdot 12t) + 0.1 \cdot \sin(2\pi \cdot 30t)$$

If explicit formula is known.

1 step. Find the highest frequency present in the signal

All frequencies: 5 Hz, 12 Hz, 30 Hz
The highest frequency is 30 Hz.

Shannon Nyquist Sampling Theorem



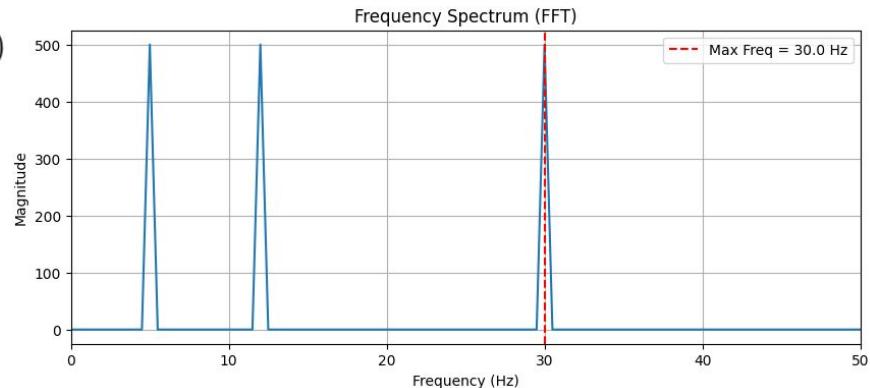
$$x(t) = 0.5 \cdot \sin(2\pi \cdot 5t) + 0.5 \cdot \sin(2\pi \cdot 12t) + 0.5 \cdot \sin(2\pi \cdot 30t)$$

1 step. Find the highest frequency present in the signal

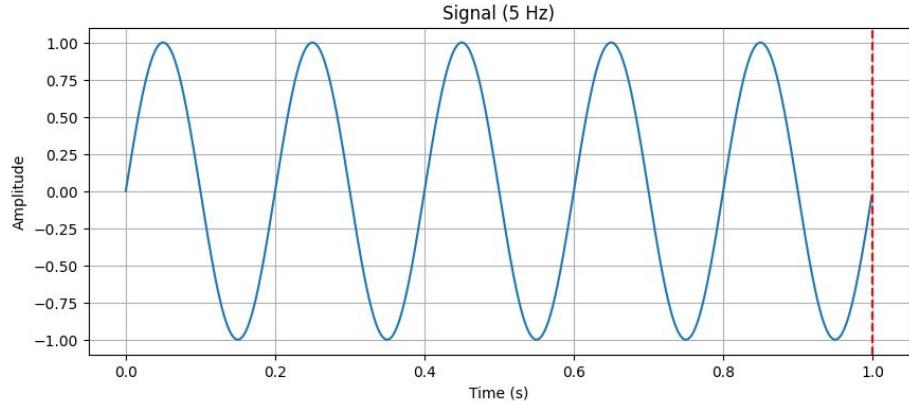
If the signal's equation is **known**.

All frequencies: 5 Hz, 12 Hz, 30 Hz.
The highest frequency is 30 Hz.

If the signal's equation is **unknown**?

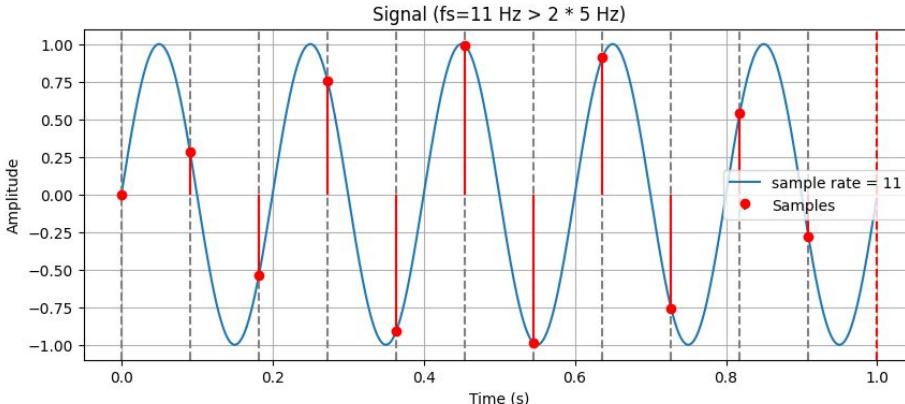


Shannon Nyquist Sampling Theorem



1 step. Find the highest frequency present in the signal

5 Hz - the highest frequency.



2 step. sample at least twice the highest frequency present in the signal.

$$f_s \geq 2 \times f_{\text{ax}}$$

Shannon Nyquist Sampling Theorem

How often should we sample to perfectly reconstruct a signal?

Key Principle: To perfectly reconstruct a signal from its samples, you must sample at least **twice the highest frequency** present in the signal.

$$f_s \geq 2 \times f_{\text{ax}}$$

- f_s = sampling frequency
- f_{ax} = highest frequency in the signal

Итак, наиболее популярные частоты сэмплирования и почему именно они?

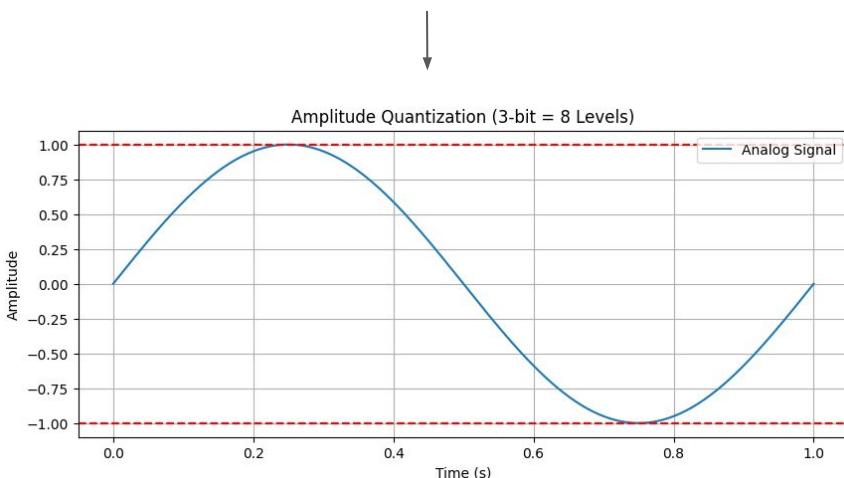
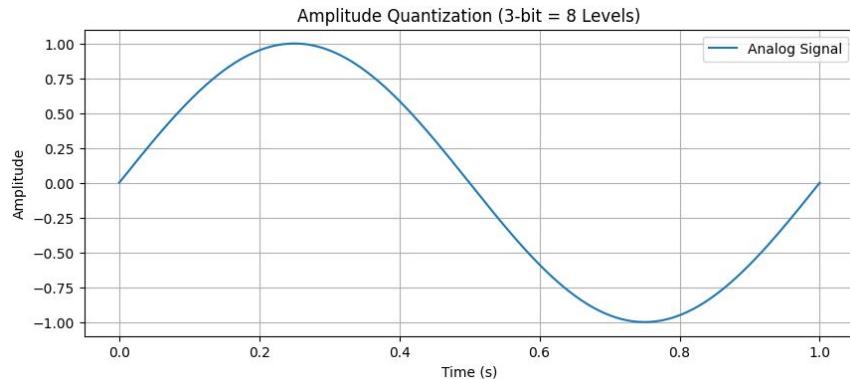
8 кГц - Если сэмплировать с частотой 8 кГц, максимальная частота, которую можно точно захватить, составляет 4 кГц. Это значит, что частоты выше 4 кГц будут потеряны или произойдет aliasing (отобразятся как более низкие частоты). Да, человеческая речь обычно звучит в диапазоне от 80 Гц до 8 кГц, но дело в том, что наиболее важные частоты для понимания речи находятся в диапазоне 300 Гц - 3,4 кГц. Согласные (например, [с, т, к]) имеют более высокие частоты (~3-8 кГц), но гласные (~300 Гц - 2 кГц) несут больший вклад в понимание речи. В телефонных системах традиционно используется дискретизация 8 кГц, так как им достаточно, чтобы речь была просто разборчивой.

16 кГц - Слышимые частоты в человеческой речи ниже 8 кГц, поэтому достаточно дискретизации речи с частотой 16 кГц. Поэтому это часто используемое значение частоты.

44 кГц - Человек может слышать частоты до ~20 кГц. Для идеального захвата 20 кГц минимальная частота должна быть 40 кГц. Берется частота 44,1 кГц (немного выше 40 кГц), чтобы избежать алиасинга + учесть несовершенство анти-алиасинговых фильтров.

Заметки полукровки.

Analog to Digital Converter: Amplitude Discretization



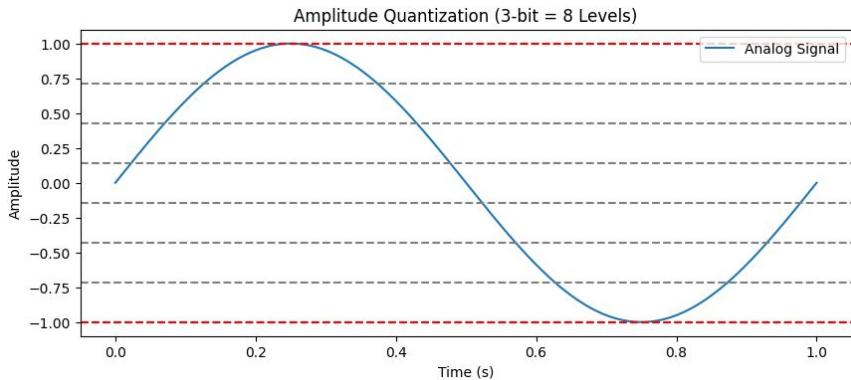
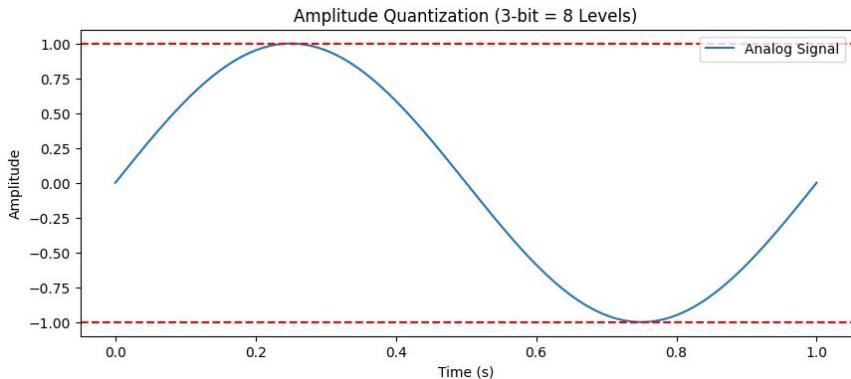
The **bit depth** of the sample determines with how much precision this amplitude value can be described.

Let's quantize amplitude in **3-bit bit depth**.

Number of levels = $2^{\text{Bit Depth}}$

Bit depth is 3-bit. It means that we have $2^3 = 8$ levels.

Analog to Digital Converter: Amplitude Discretization



Let's quantize amplitude in **3-bit bit depth**.

Bit depth is 3-bit. It means that we have $2^3 = \textbf{8 levels}$
[+ 1 (include zero) = 9 levels]



Amplitude range: [-1.0, 1.0]

discretize to 8 levels

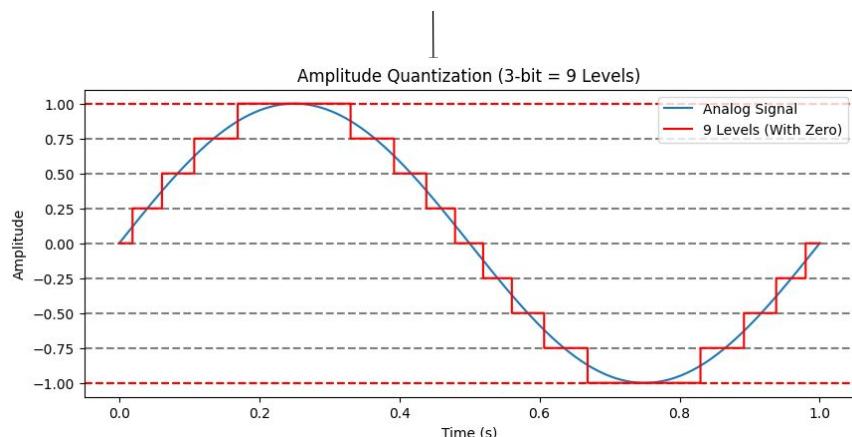
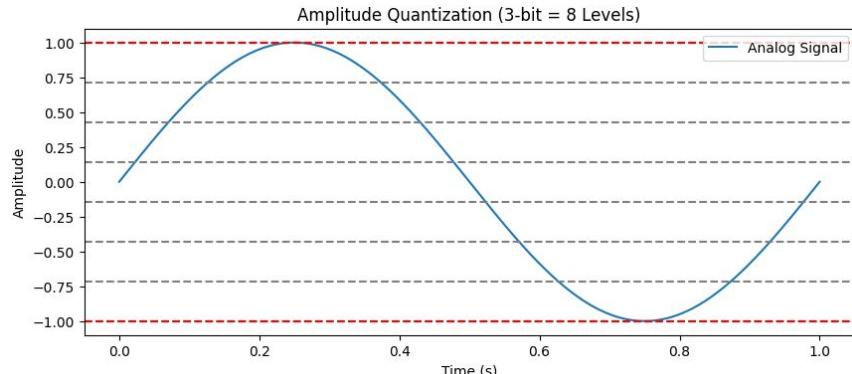
```
print(np.linspace(-1, 1, 8))
```



[-1.0, -0.714, -0.429, -0.143, 0.143, 0.429, 0.714, 1.0]

continuous values divided to 8 discrete values

Analog to Digital Converter: Amplitude Discretization



Let's quantize amplitude in **3-bit bit depth**.

Bit depth is 3-bit. It means that we have $2^3 = \mathbf{8 \text{ levels}}$
+ 1 (include zero) = 9 levels.



Amplitude range: [-1.0, 1.0]



discretize to 9 levels

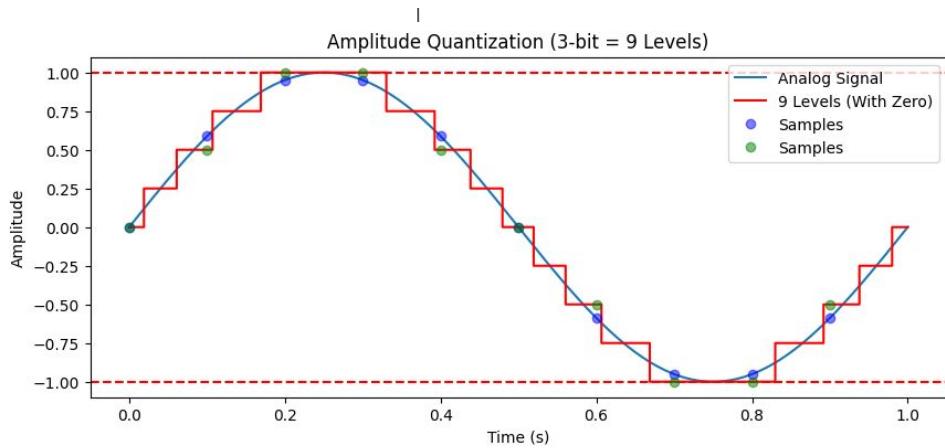
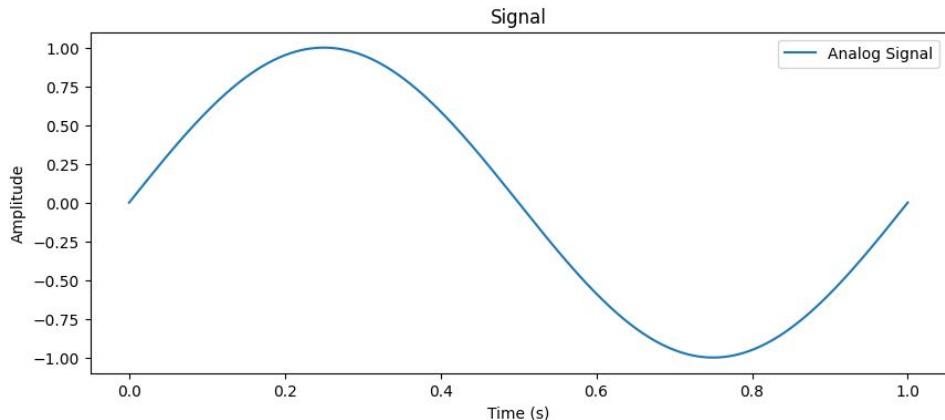
```
print(np.linspace(-1, 1, 8))
```



```
[-1. -0.75 -0.5 -0.25 0. 0.25 0.5 0.75 1. ]
```

continuous values divided to 9 discrete values

Analog to Digital Converter: Amplitude Discretization



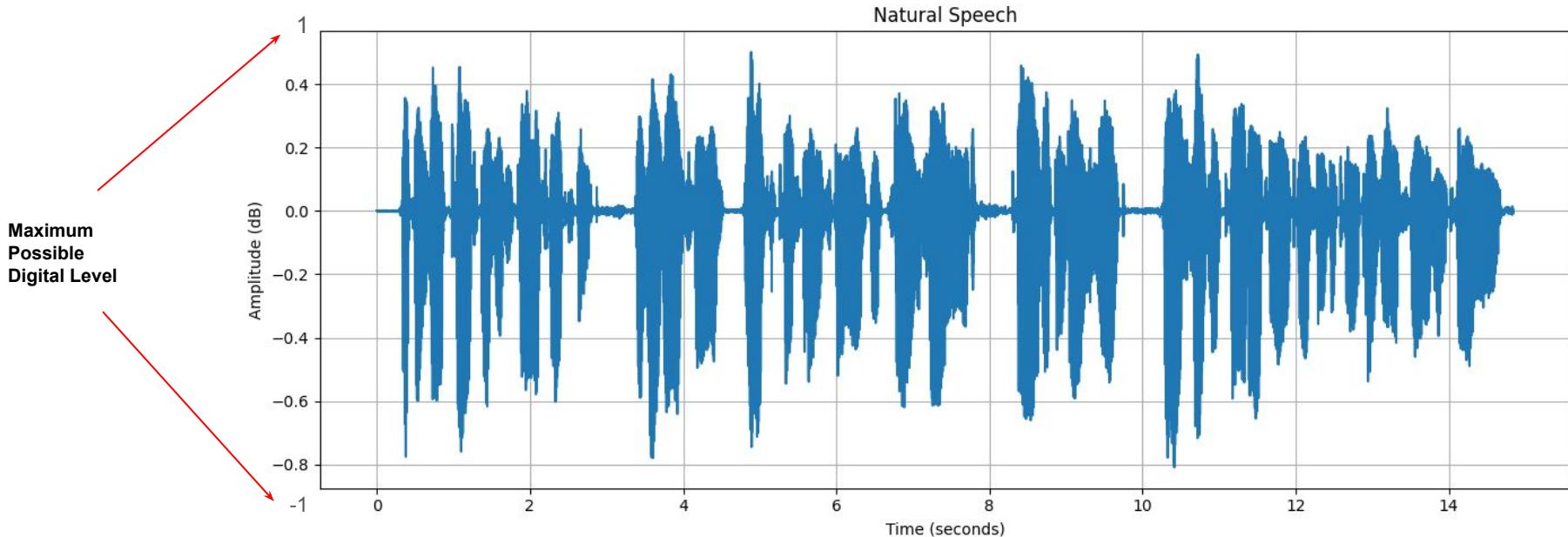
Наиболее частые значения bit depths это **16-bit** и **24-bit**. Поскольку квантование подразумевает округление непрерывного значения до дискретного, процесс дискретизации вносит шум. Чем выше bit depths, тем меньше этот **quantization noise**. На практике **quantization noise 16-bit** аудио уже достаточно мал, чтобы быть слышимым, и использование больше bit depths обычно не требуется.

Вы также можете встретить 32-битное аудио. В нем выборки хранятся в виде значений с плавающей точкой, тогда как в 16- и 24-битном аудио используются целочисленные выборки. Точность 32-битного значения с плавающей точкой составляет 24 бита, что дает такую же битовую глубину, как и у 24-битного аудио.

Ожидается, что образцы звука с плавающей точкой будут лежать в диапазоне [-1.0, 1.0].

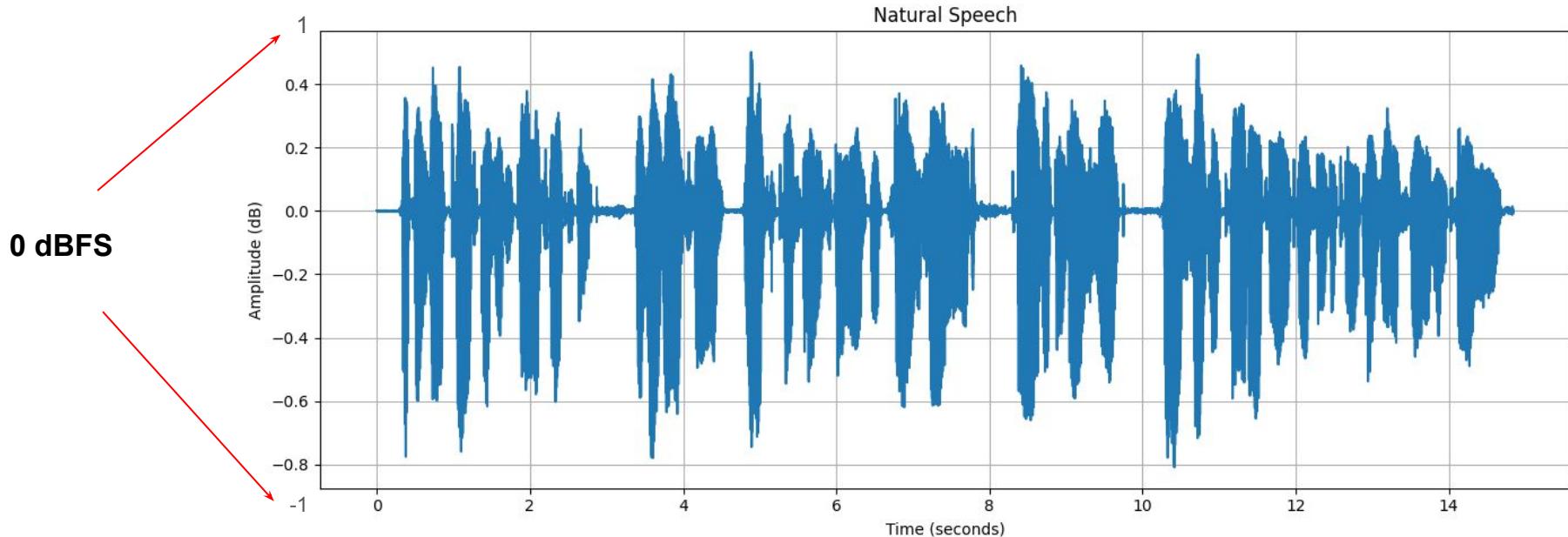
Заметки полукровки.

Properties of Sound Waves: Real-world sound vs Digital sound dB scale



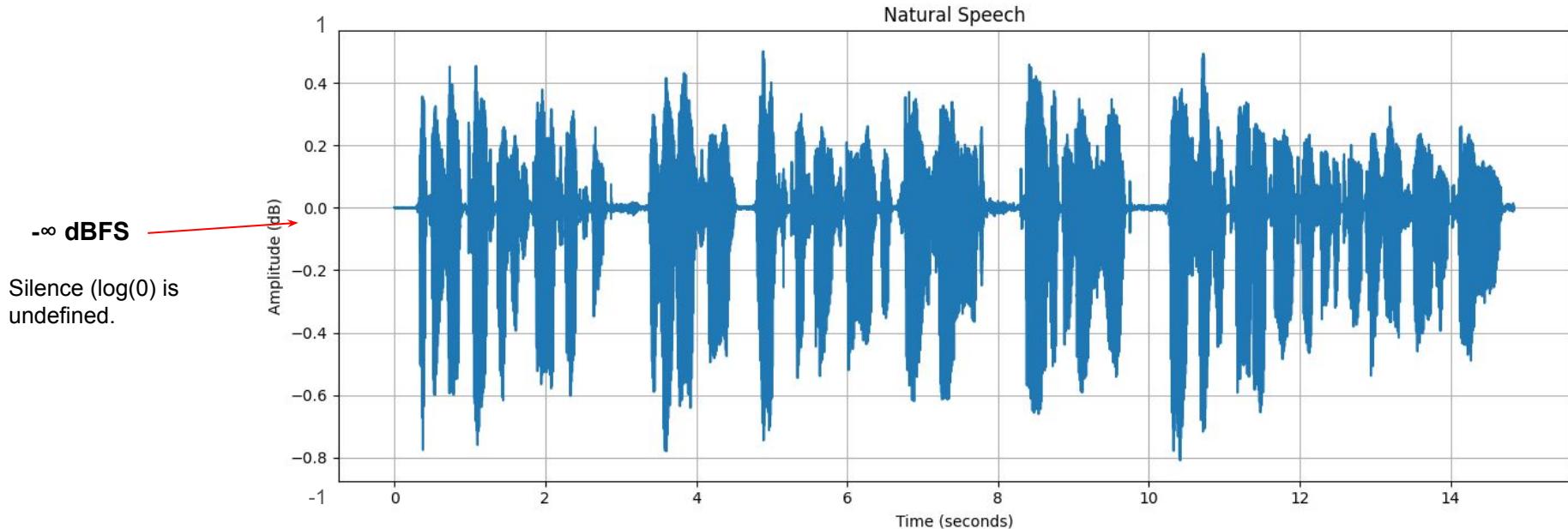
However, **for digital audio signals, 0 dB is the loudest possible amplitude**, while all other amplitudes are negative. 0 dBFS (Decibels Relative to Full Scale) is defined as the **maximum possible digital level** (our +1 or -1 in the normalized PCM data).

Properties of Sound Waves: Real-world sound vs Digital sound dB scale



However, **for digital audio signals, 0 dB is the loudest possible amplitude**, while all other amplitudes are negative. 0 dBFS (Decibels Relative to Full Scale) is defined as the **maximum possible digital level** (our +1 or -1 in the normalized PCM data).

Properties of Sound Waves: Real-world sound vs Digital sound dB scale



However, **for digital audio signals, 0 dB is the loudest possible amplitude**, while all other amplitudes are negative. 0 dBFS (Decibels Relative to Full Scale) is defined as the **maximum possible digital level** (our +1 or -1 in the normalized PCM data).

Number of channels

Lecture: Intro to Speech
Processing



Number of channels: Mono, Stereo, Spatial Audio



Здесь важны две вещи: каким образом вы записываете, и как человек будет слушать это и на каком устройстве.

Audio Formats

Lecture: Intro to Speech
Processing



Audio formats: mono, stereo

Эти аудиоформаты как раз хранят оригинальные дискретованные аудиосигналы.

- Uncompressed:

- **wav** (on Windows),
- **aiff** (on MacOS),
- **au** (on Unix)

Чаще всего дамасеты хранятся в этих аудио форматах.

Уменьшение размера в два раза.

- Lossless compression (2:1):

- **flac** (Free Lossless Audio Codec),
- **alac** (Apple Lossless Audio Codec)

- Lossy compression:

- **MP3**,
- **Opus** (Discord, WhatsApp)
- **AAC** (iTunes Music Store)

Уменьшение размера за счет удаления информации неслышимой и не различимой человеком.



https://en.wikipedia.org/wiki/Audio_file_format

Часто пишут следующую фразу “wav and aiff formats can store LPCM”. Что это значит?

PCM (pulse code modulation) это как раз метод дискретизации аналогового сигнала. LPCM (Linear pulse code modulation) это один из видов PCM, где уровни квантации линейно равномерны. Таким образом “can store LPCM” означает, что этот формат может хранить дискретованные оригинальные аудиосигналы.

Заметки полукровки.

Audio formats: spatial

Format	Type	Key Strengths	Common Applications
Dolby Atmos	Object-based	Dynamic object placement; scalable rendering	Movies, music, gaming
MPEG-H 3D Audio	Hybrid	Open standard; supports objects/channels/Ambisonics	Broadcasting, music streaming
Sony 360RA	Object-based	Music-focused immersion; personalized HRTF	Music streaming, PlayStation
Ambisonics	Scene-based	360° capture; ideal for VR/AR	360° video, field recordings
Apple Spatial Audio	Proprietary	Head tracking; seamless Apple integration	Apple Music, FaceTime, visionOS
Mach1 Spatial	Channel-based	Format conversion without metadata	Developer tools, game audio
Game Engine Objects	Runtime object-based	Interactive sound placement	Gaming, VR simulations



Yermekova Assel

Telegram: @zametki_polukrovki - тут я рассказываю какие то вещи.

Спасибо за внимание!

Если у вас остались вопросы, пишите
в чат курса.