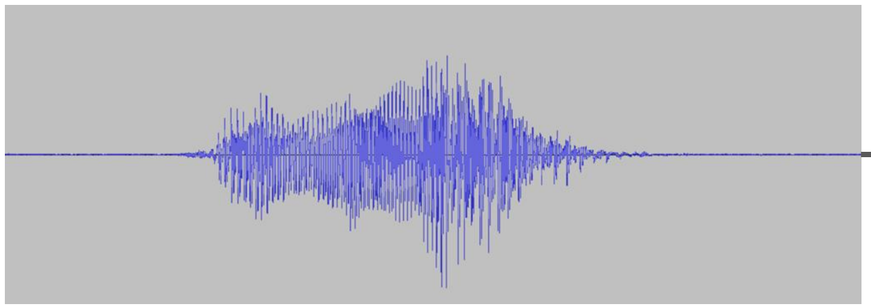
1 second audio sample

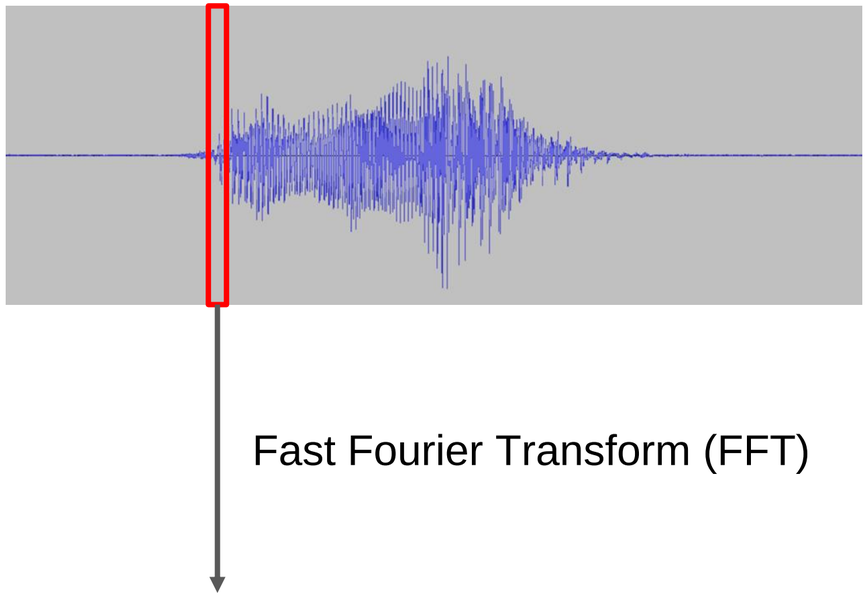


Training

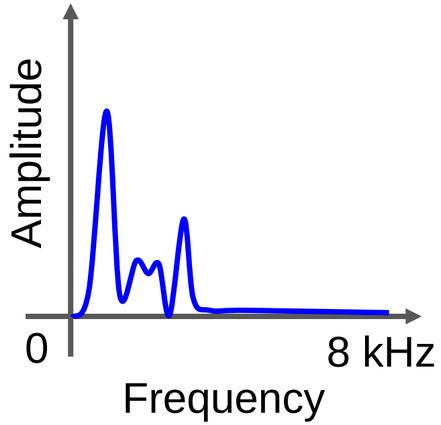
Model

Fourier Transform

1 second audio sample (“hello”)



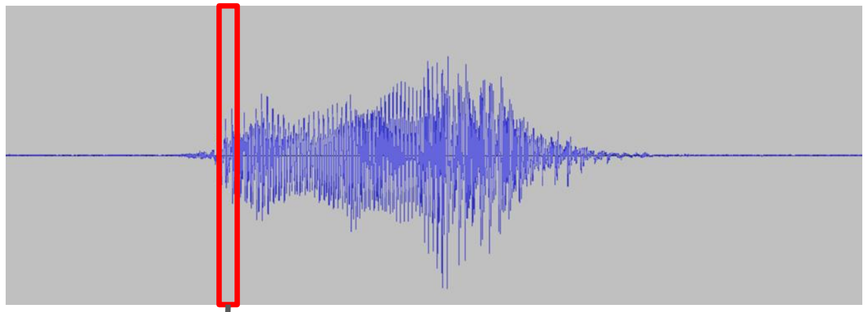
Fast Fourier Transform (FFT)



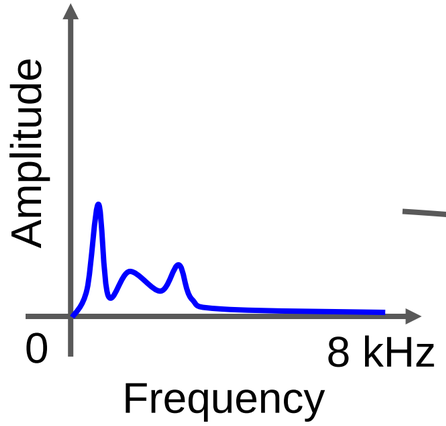
Voice frequency range: 300 - 3400 Hz

Fourier Transform

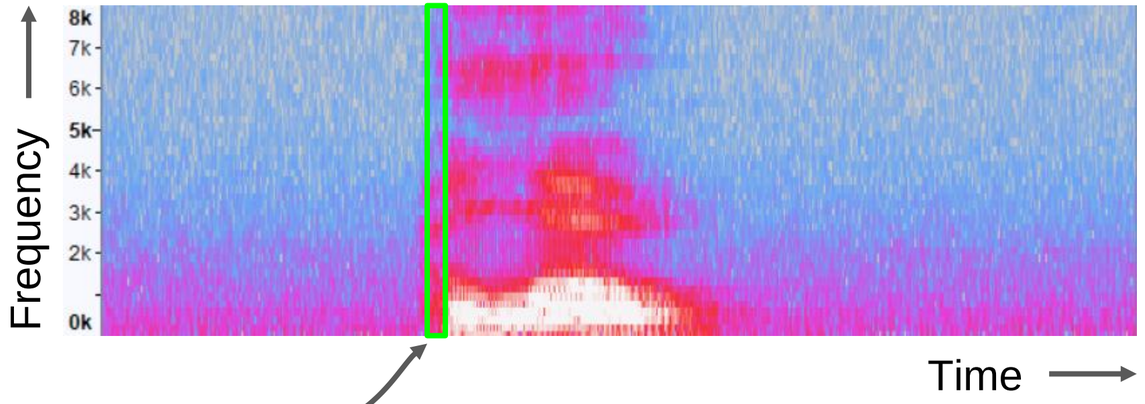
1 second audio sample (“hello”)



Fast Fourier Transform (FFT)



Spectrogram

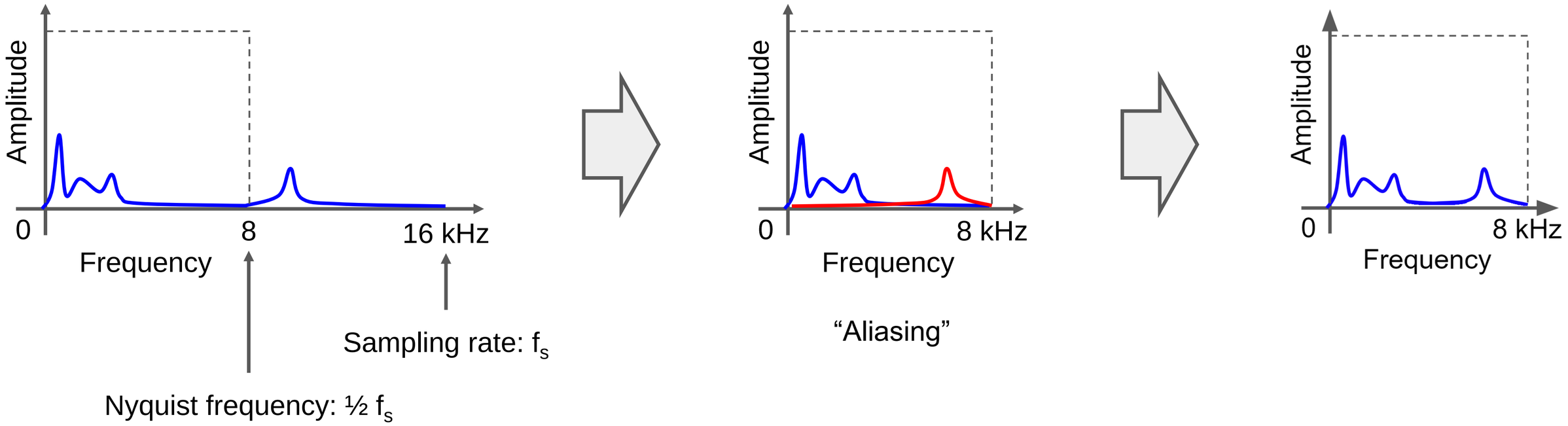


Voice frequency range: 300 - 3400 Hz

Nyquist-Shannon Sampling Theorem

$$f_s > 2B$$

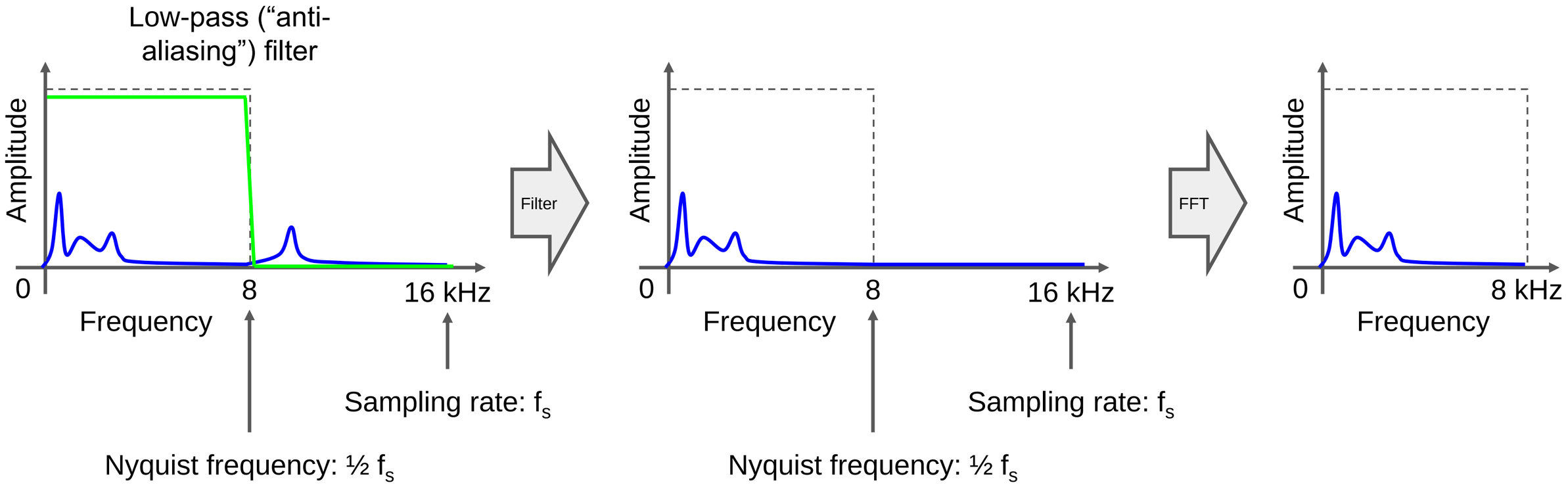
f_s is the sampling frequency (Hz)
B is the highest frequency component (Hz)



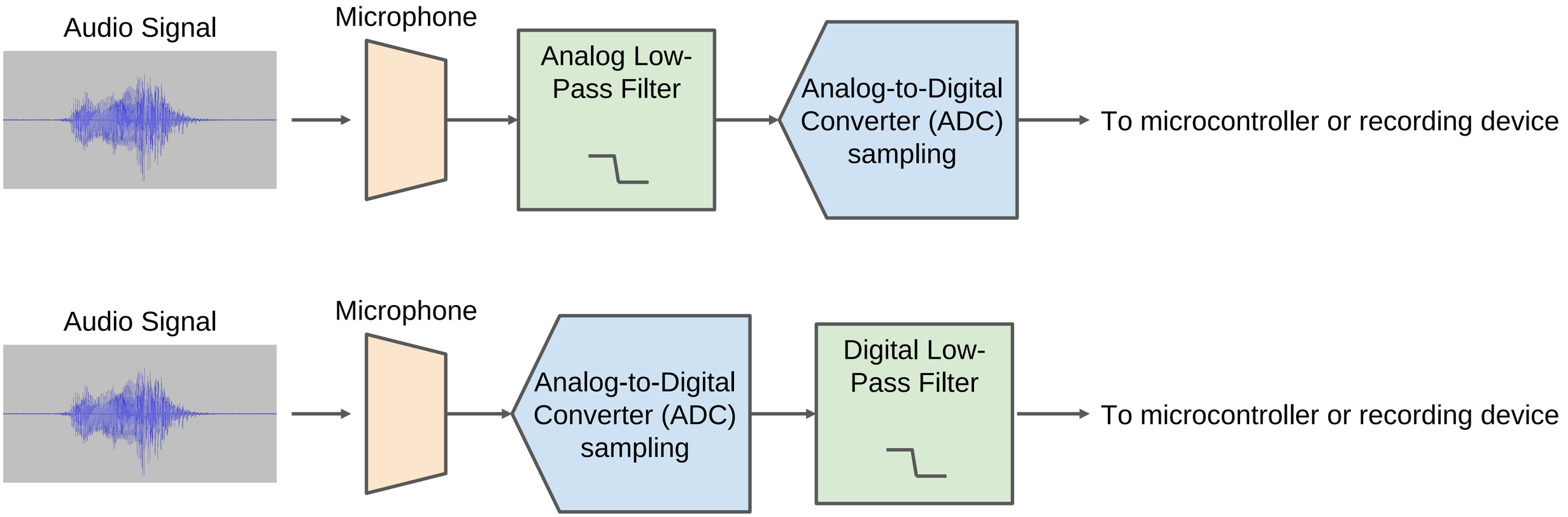
Nyquist-Shannon Sampling Theorem

$$f_s > 2B$$

f_s is the sampling frequency (Hz)
B is the highest frequency component (Hz)



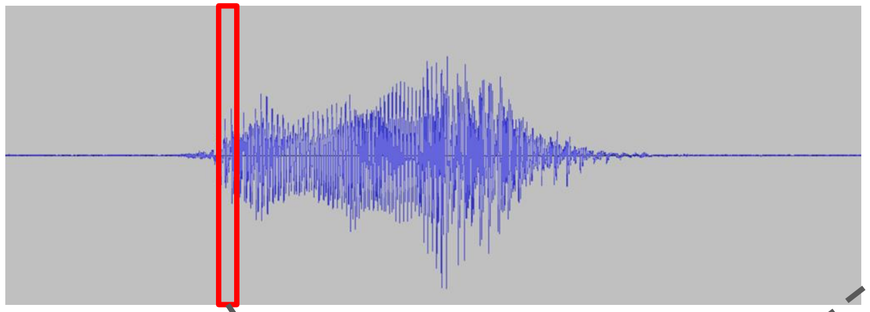
Nyquist-Shannon Sampling Theorem



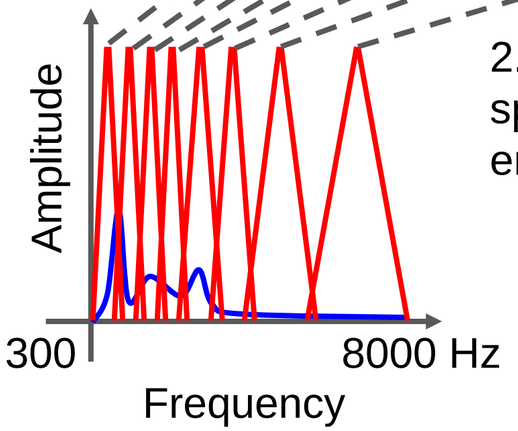
Nyquist-Shannon Sampling Theorem

Lesson: make sure what you're recording is less than half the sampling frequency!

Mel Frequency Cepstral Coefficients (MFCCs)



1. Take FFT of window



2. Compute Mel-spaced filterbank energies

1.063	Low freq. energy
0.407	
0.229	
0.164	
0.352	
0.015	
0.012	
0.007	High freq. energy

log()

3. Compute log of each value

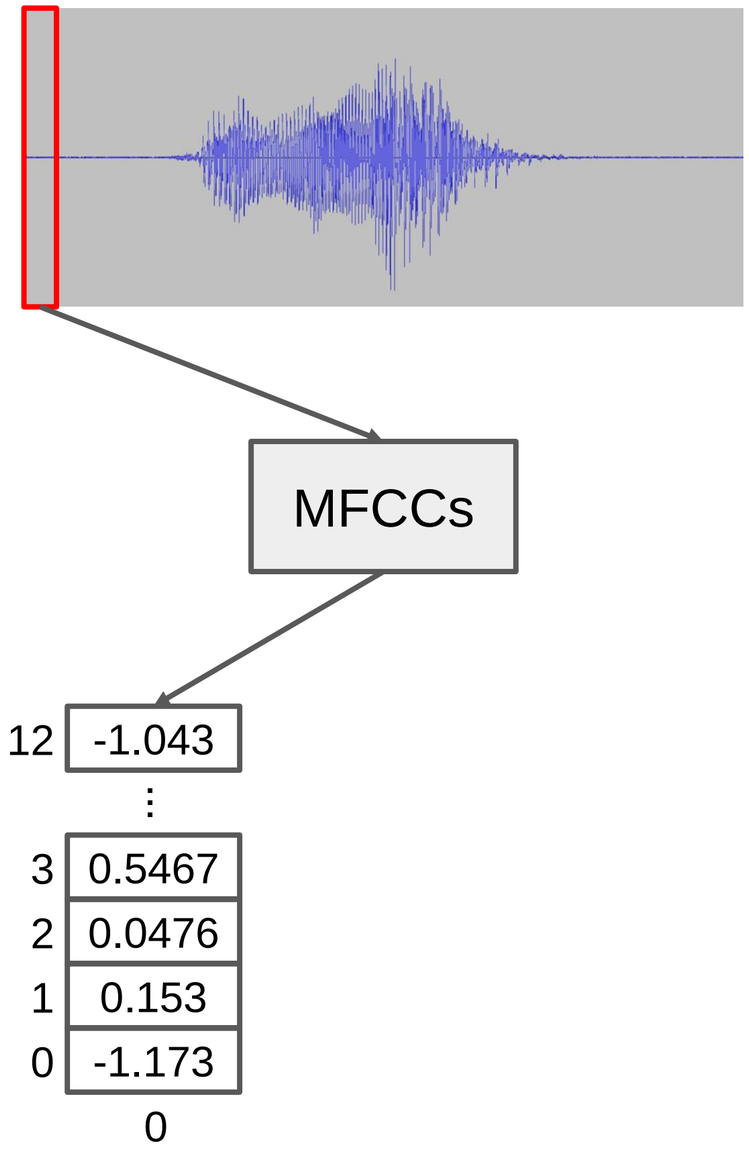
4. Compute Discrete Cosine Transform

DCT

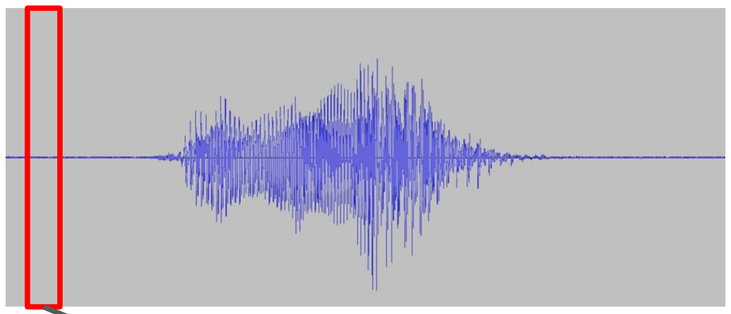
-16.38	0
4.477	1
-5.061	2
-0.749	3
-2.340	4
~~-6.093~~	5
~~-1.043~~	6
~~3.221~~	7

Mel Frequency Cepstral Coefficients

Mel Frequency Cepstral Coefficients (MFCCs)



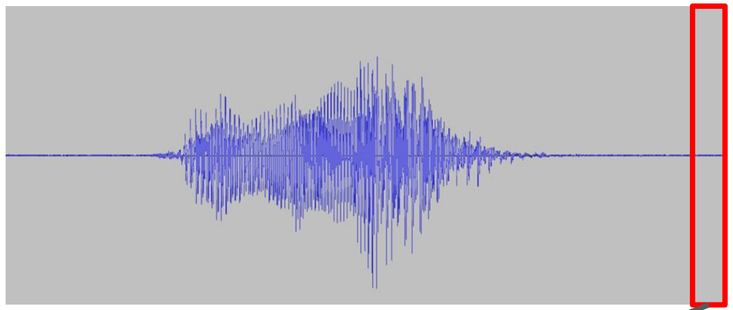
Mel Frequency Cepstral Coefficients (MFCCs)



MFCCs

	0	1
12	-1.043	-0.816
⋮	⋮	⋮
3	0.5467	0.442
2	0.0476	0.836
1	0.153	-0.671
0	-1.173	0.462

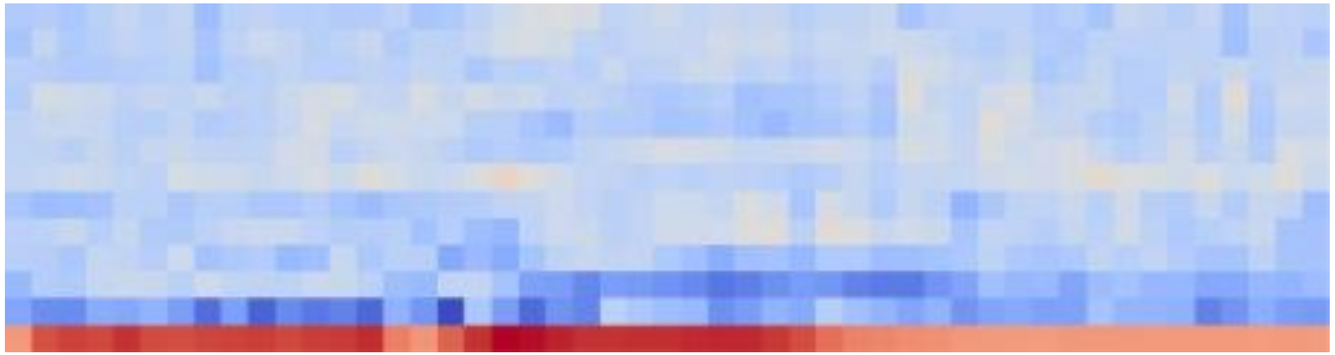
Mel Frequency Cepstral Coefficients (MFCCs)



MFCCs

12	-1.043	-0.816	···	-0.184
	⋮	⋮		⋮
3	0.5467	0.442	···	-0.523
2	0.0476	0.836	···	0.185
1	0.153	-0.671	···	-0.248
0	-1.173	0.462	···	-1.218
	0	1		48

“stop”



“hello”