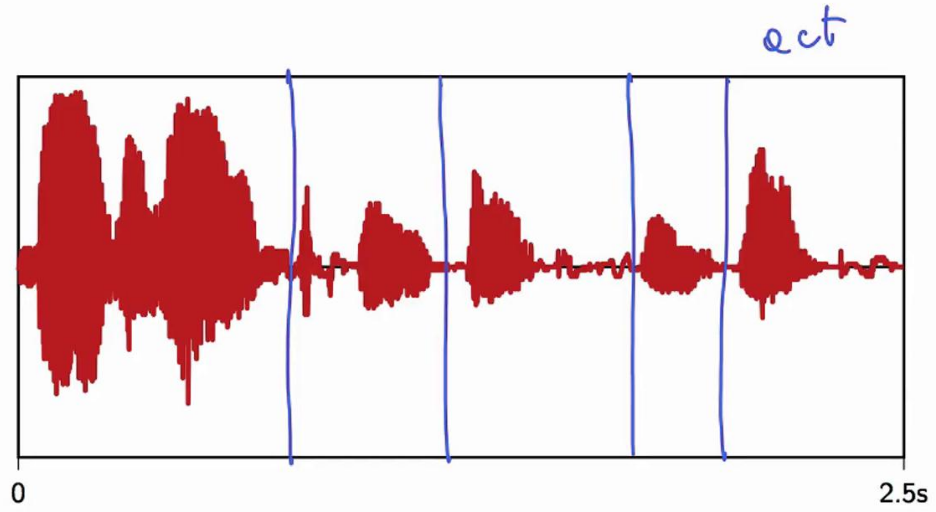
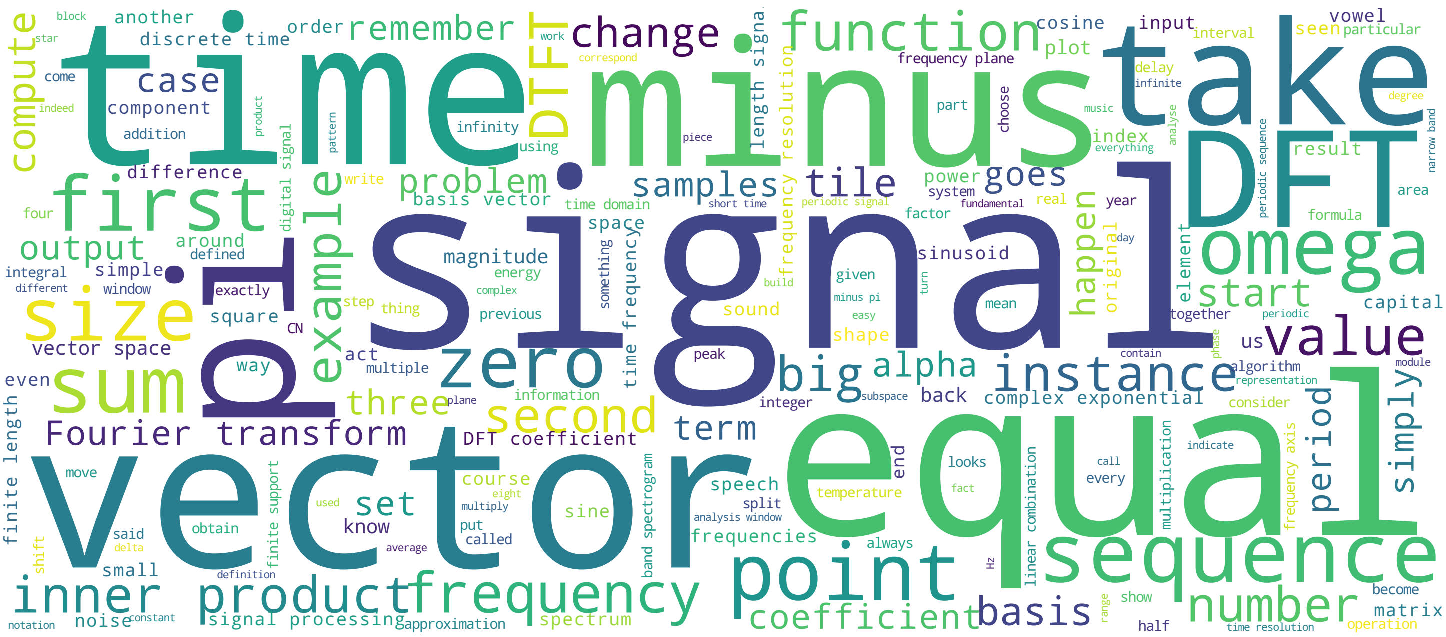
Speech analysis

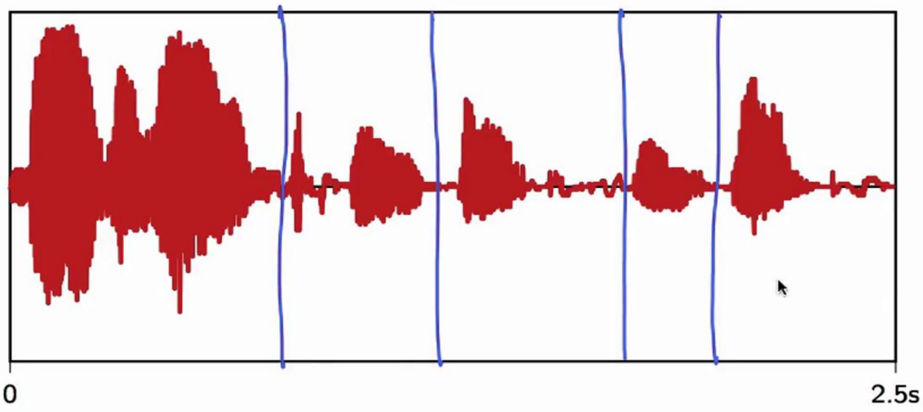




Search MOOC

Video

Speech analysis

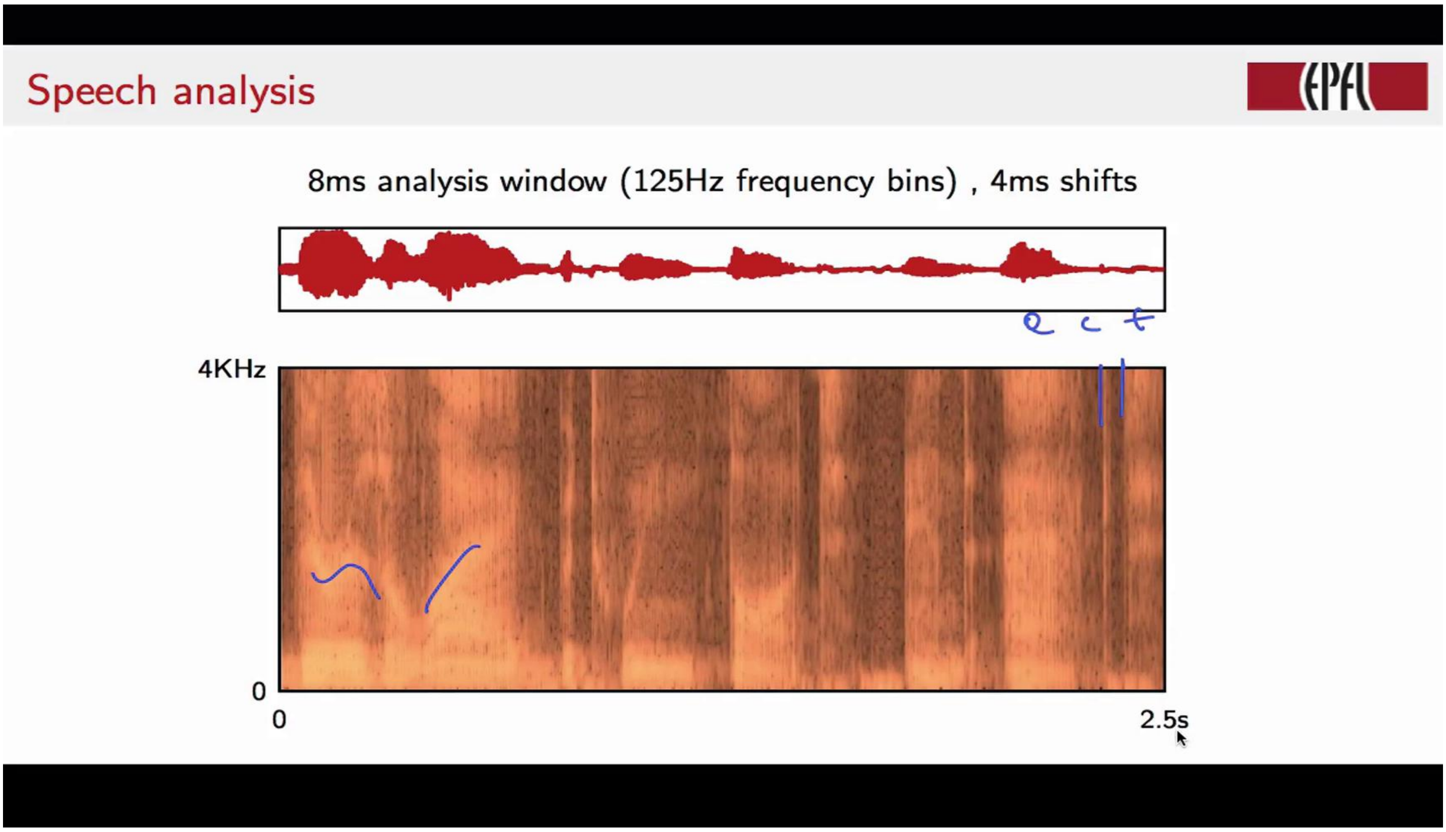


Spectrograms are particularly useful and particularly popular in speech analysis. Speech is a particularly difficult signal to analyse because the mechanism of speech production alternates between wildly different modes of operation. When you pronounce a vowel sound, like a,for instance, you're producing a harmonic sound that resonates in your body and that contains a very well structured harmonic content. On the other hand, consonants have a noiselike structure mostly, and their spectral makeup is completely different. So it would be futile to try and come up with a global spectral representation for an entire speech utterance. We need to split the speech into pieces and analyse the species in sequence. Here's an example. This is a sentence from a speech corpus that is used in speech analysis algorithms. There is a lag between thought and act. So in the time domain, the waveform can be split like this. This part here is there is a lag, this is between thought and act. Here we are in the presence of a portion of speech that is rich in vowels and here we have single words that have both vowels and consonants in particular. For instance, look at this, this is act.

Notes

Summary

0m 00s



We have A and then C and T, which are noiselike pulses. The speech signal was sampled at eight khz. Let's try and see what wide band spectrogram can tell us about the structure of this utterance. If we take an eight milliseconds analysis window, which gives us frequency bins of 125 Hz, which is quite wide, we get a plot like this. We cannot see much in terms of frequency resolution, although we start to see some patterns here that will be clearer later. What we do have is very precise onsets for the consonants. Remember, this is act. See here, you see the beginningof the C, and here you see the beginning of T in this onset of a highenergy band in the spectrogram. The same could be said for the other consonants.

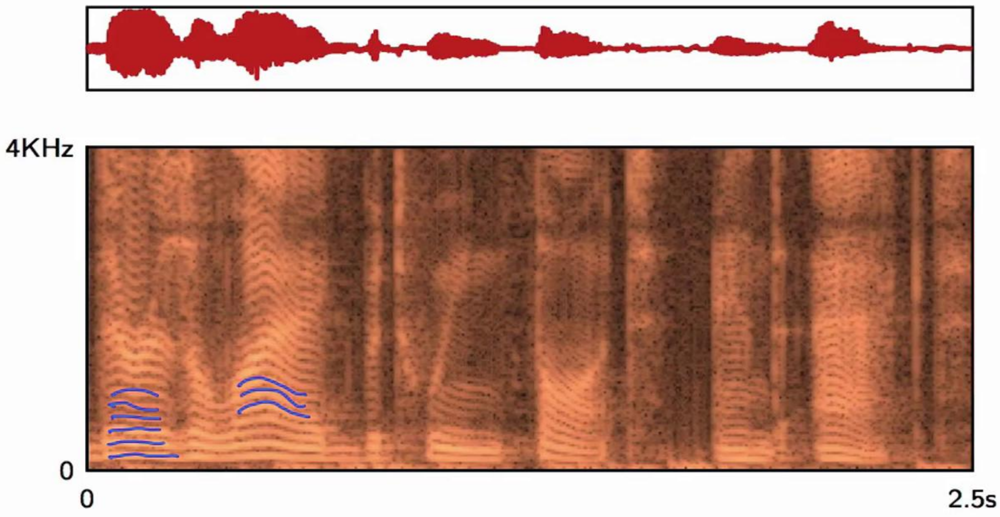
Notes

Summary

1m 28s

Speech analysis

32ms analysis window (31Hz frequency bins), 4ms shifts

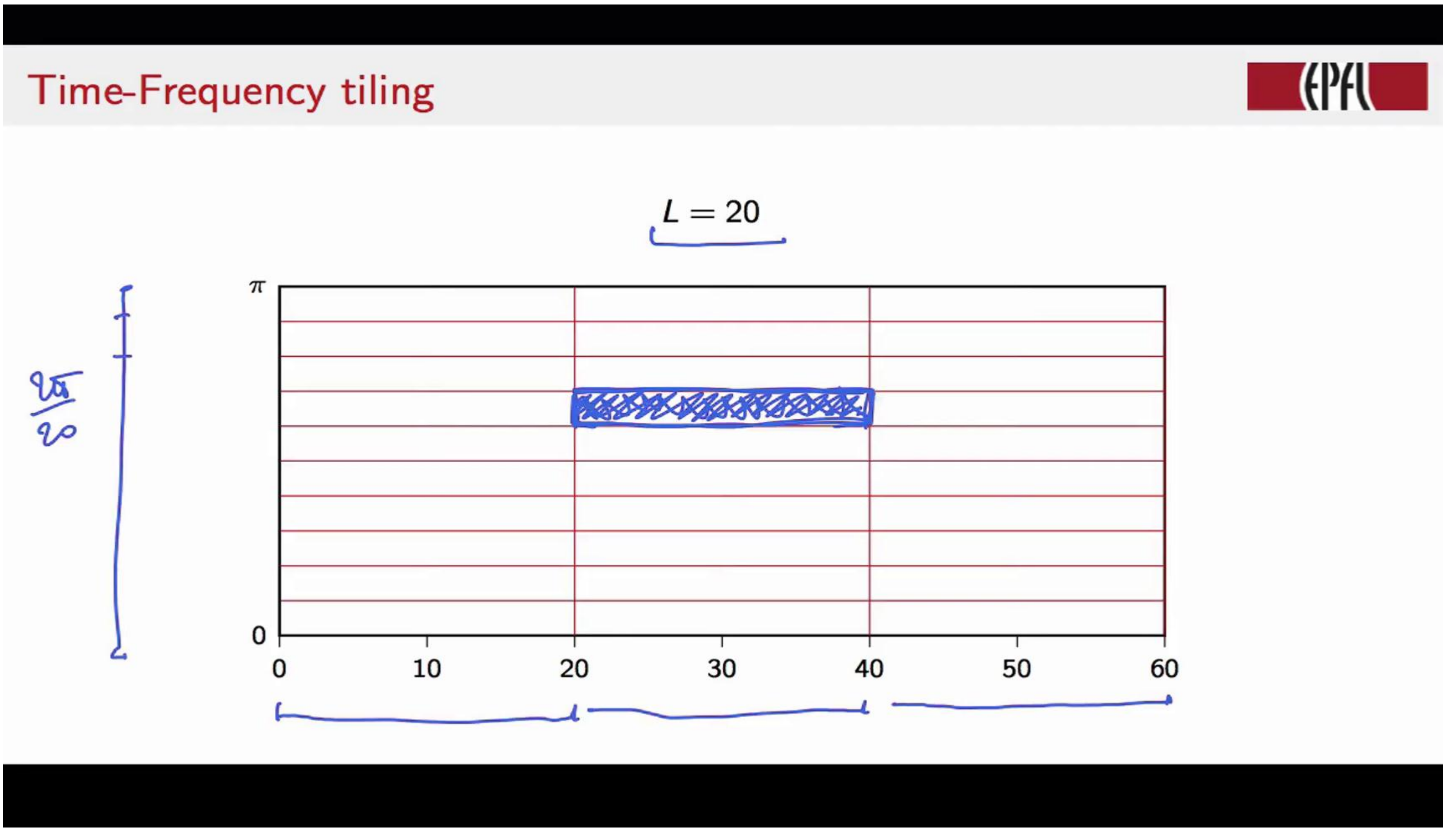


If we use a narrow band spectrogram, so we increase the analysis window to 32 milliseconds, then we have a resolution of 31 Hz. Here we see the harmonic structure of speech. Here the focus is on the vowels, and here you can see that each vowel contains a harmonic structure that depends on the pitch of the vowel. It will change from speaker to speaker, from male speaker to female speakers. You can also see how in speaking we modulate and change the frequency of the vowels to give a certain intonation to our utterance. Narrow band spectrogram will give us information on the harmonic parts of speech. Wideband spectrogram will give us information on the pulselike and noiselike consonant sounds in speech.

Notes

Summary

2m 22s

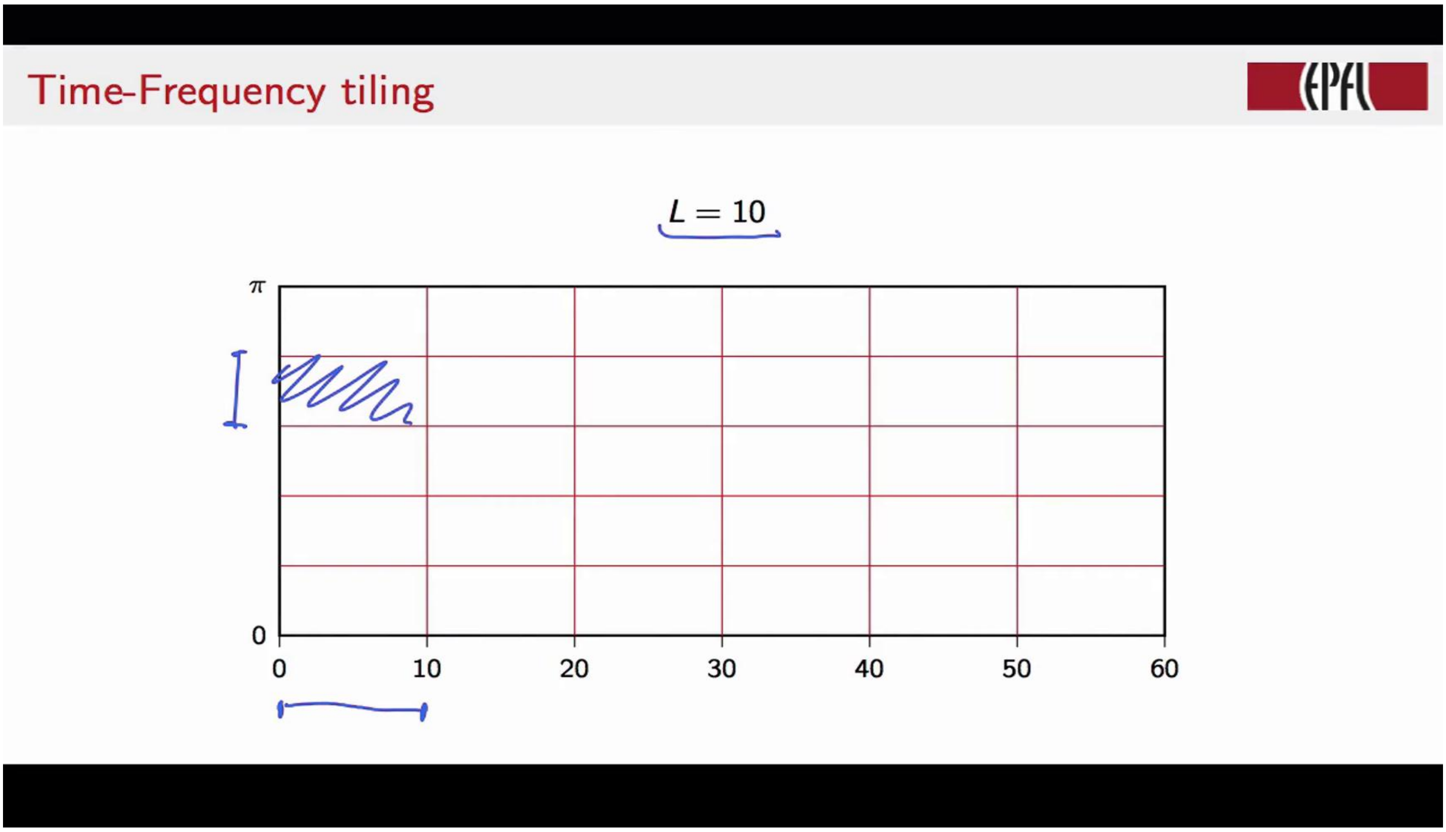


The short time for your transform determines a tiling of the time frequency plane, where the size of each tile is specified by the time and the frequency resolution of the STF. Suppose you choose a window size equal to 20. What we have is a subdivision of the time axis into chunks that are 20 samples long and a subdivision of the frequency axis into bins, each one of which is two pi over 20. So each tile in the frequency plane will have a horizontal size of 20 and a vertical size of two pi over 20. Everything that happens in the time frequency plane within this tile will be summed up by just one STFT value.

Notes

Summary

3m 16s



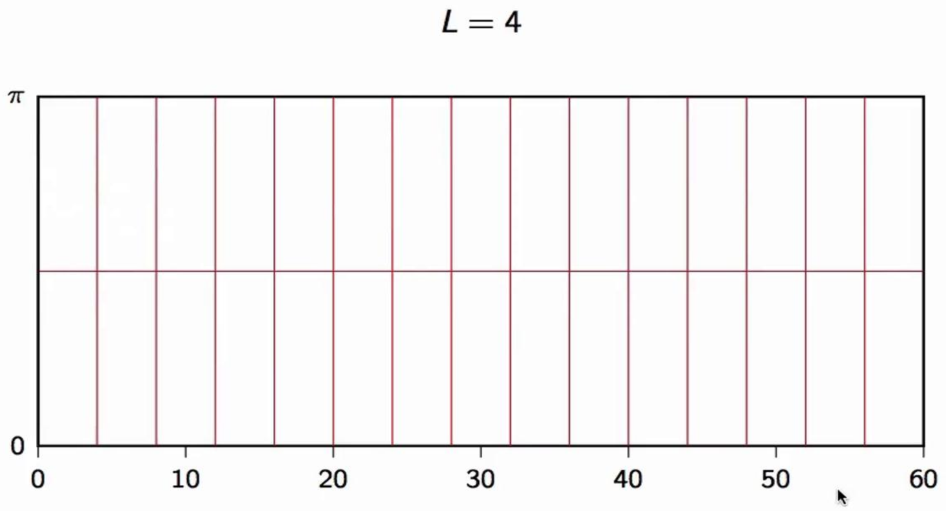
If we change the size of the window, suppose we take L equal to 10, then we narrow the size of the tile in the time axis, but we widen the size of the tile in the frequency axis. Although the shape of the tiles change, the number of tiles remains the same, because the area of each tile remains constant.

Notes

Summary

4m 03s

Time-Frequency tiling



Similarly, if we shorten the window even more, we have a different arrangement of the tiles, but the size of each tile remains the same.

Notes

Summary

4m 27s

Food for thought

- time "resolution" $\Delta t = L$
- frequency "resolution" $\Delta f = 2\pi / L$
- $\Delta t \Delta f = 2\pi$

uncertainty principle!

This is actually quite self-evident. If the time resolution is L, the frequency resolution is two pi over L, and therefore the product, namely the area of each tile is the constant two pi. In a nutshell, this is the uncertainty principle in time frequency analysis. It states that we cannot arbitrarily narrow our focus both in time and in frequency. If we want a higher time resolution, we will necessarily have to give up frequency resolution and vice versa.

Notes

Summary

4m 36s

Even more food for thought

more sophisticated tilings of the time-frequency planes
can be obtained with the *wavelet* transform

A short time Fourier transform leads to a very simple uniform tiling of the time frequency plane and more sophisticated structures have been the subject of research and in particular of a branch of signal processing called wavelet analysis. For those of you who are curious about wavelets, we recommend you check out the links that we provide in the bibliography for the class.

Notes

Summary

5m 08s