Spraakverwerking per computer

David Weenink

Instituut voor Fonetische Wetenschappen
ACLC
Universiteit van Amsterdam

Het spectrogram

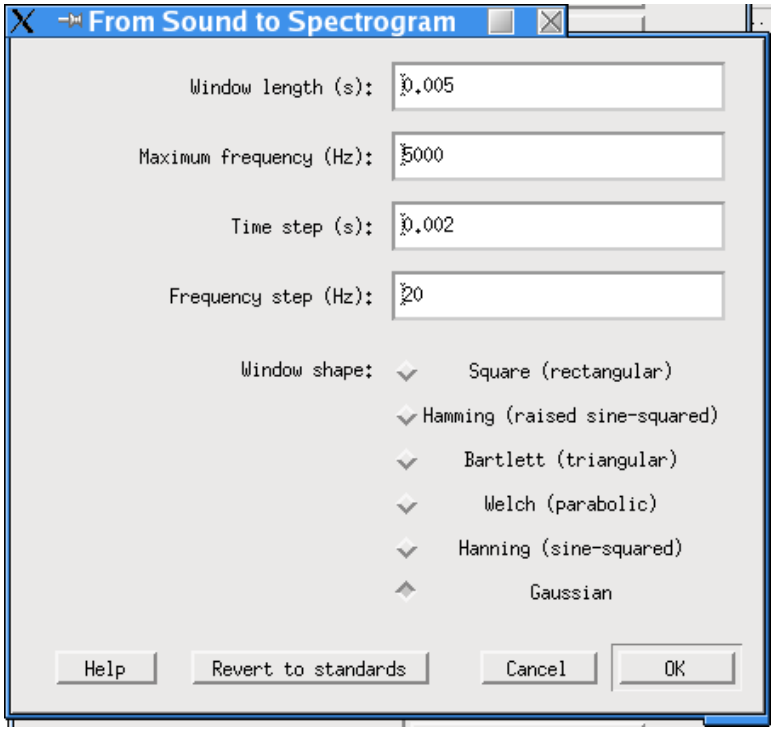
Het Spectrogram representeert een acoustische tijd-frequentie representatie van een geluid: de power spectral density $P(f, t)$, uitgedrukt in $\mathrm{Pa}^2/\mathrm{Hz}$.

Het is bemonsterd in punten die op gelijke afstanden van elkaar liggen, en gecentreerd zijn om tijdstippen t_i en frequenties f_j.

Populair: "de sterkte van frequenties als functie van de tijd". Deze "sterkte" wordt aangegeven via zwarting: hoe zwarter hoe sterker.

Een doorsnede van een analyse-object in de tijd wordt een frame genoemd.

Sound: To Spectrogram... (form)



- Help: Sound: To Spectrogram...

Sound: To Spectrogram... (window length)

Bepaalt de *duur* van het analysevenster en
daardoor ook de *bandbreedte* van spectrale analyse.

Breedband	5 ms	$B = 260$ Hz
Smalband	30 ms	$B = 43$ Hz

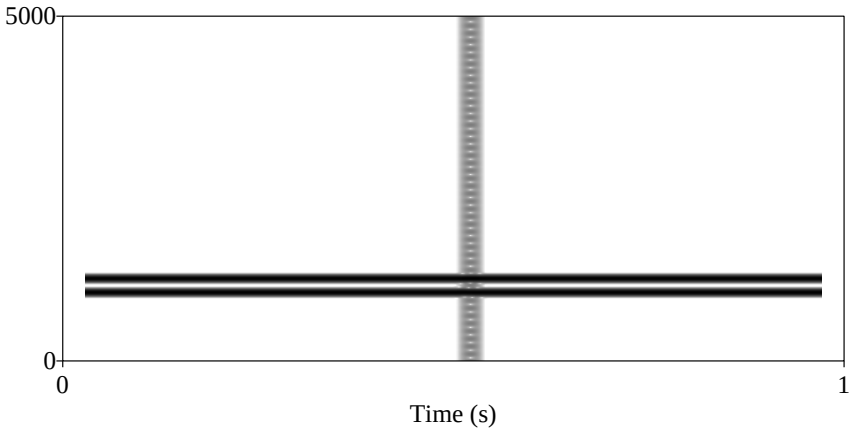
Als window length b.v. 0.005 s: $[t_m - 0.0025, t_m + 0.0025]$
t_m is midden

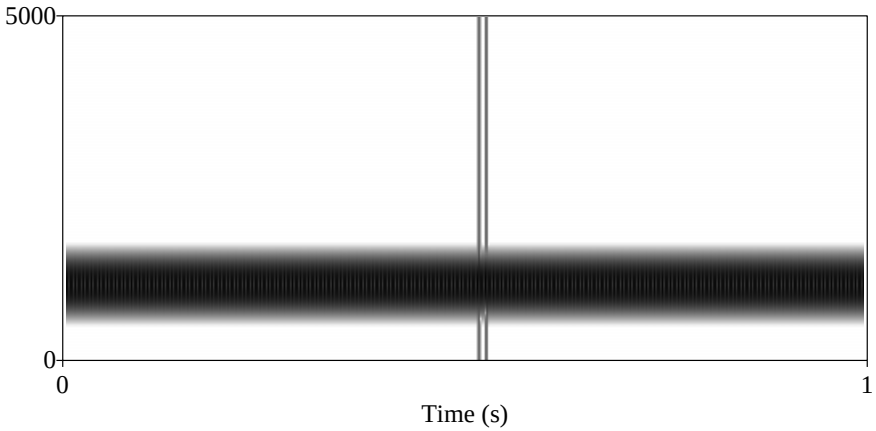
Sound: To Spectrogram... (Maximum frequency)

- Maximum frequency (Hz)
  the maximum frequency subject to analysis, e.g. 5000 Hertz. If it is higher than the Nyquist frequency of the Sound (which is half its sampling frequency), some values in the result will be zero (and will be drawn in white by Spectrogram: Paint...
- Time step (s)
  the distance between the centres of subsequent frames, e.g. 0.002 seconds. This determines the number of frames of the resulting Spectrogram. For instance, if the Sound is 1 second long, and the time step is 2 milliseconds, the Spectrogram will consist of almost 500 frames (not exactly 500, because no reliable spectrum can be measured near the beginning and end of the sound).

Smalband en breedband


```
Create Sound from formula...  bn Mono 0 1 11025
...  0.3*(sin(2*pi*1000*x) + sin(2*pi*1200*x)) +
...  (col=5700) + (col=5800)
```





- "Narrow band"
  Window length 0.030 s
- "Broad band"
  Window length 0.005 s

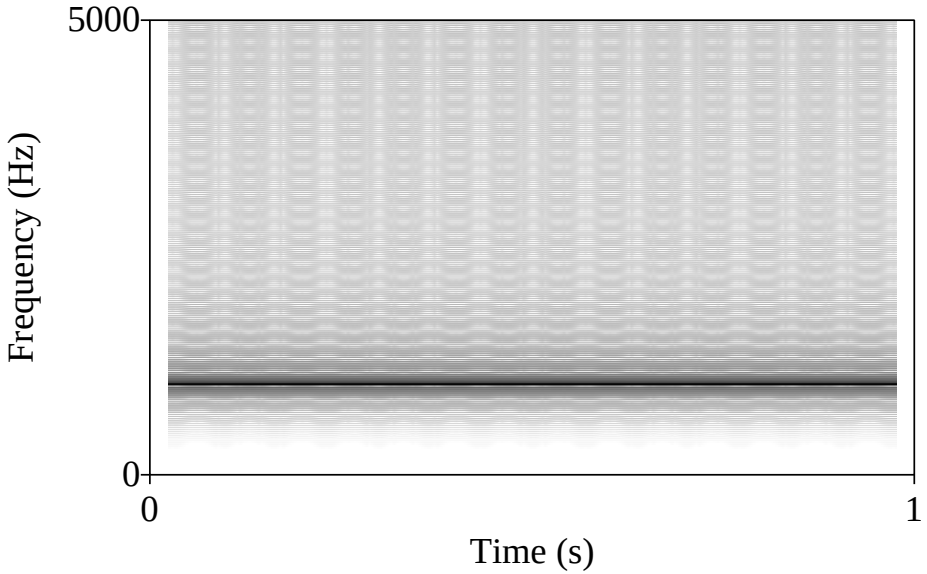
```
Paint...  0 0 0 0 100 y 50 0 0 y
```

Sound: To Spectrogram... (vensters)

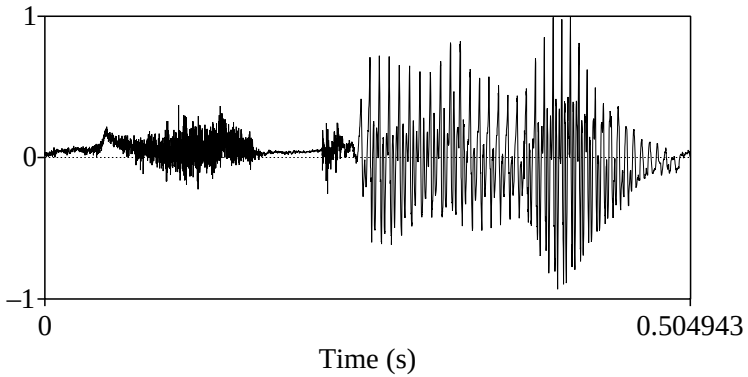
Keuze uit: Square, Hamming, Bartlett, Welch, Hanning, Gaussian

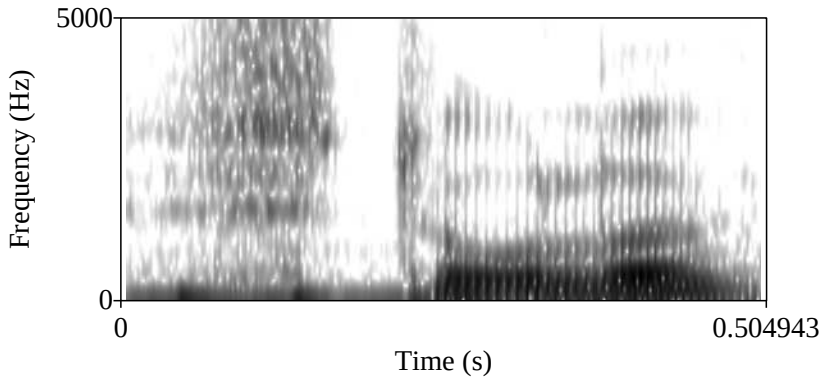
Wij gebruiken altijd de "Gaussian" omdat alle andere bij-effecten vertonen:

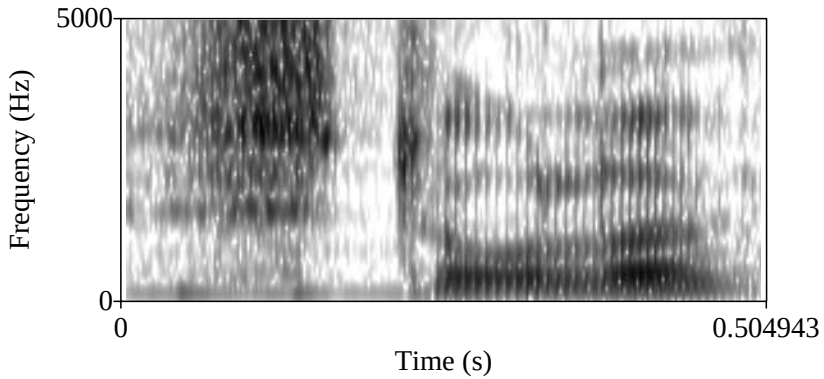
```
Create Sound from formula...  s Mono 0 1 11025 sin (2*pi*1000*x)
```



Pre-emphasis

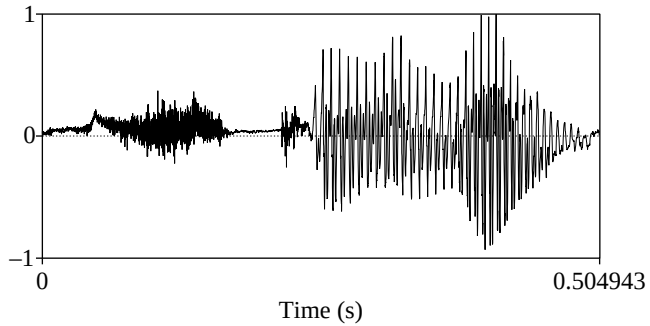


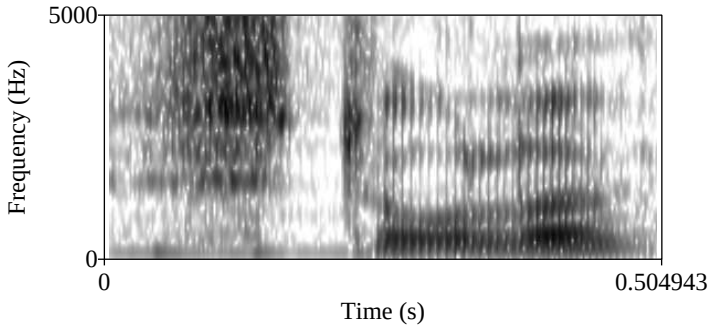


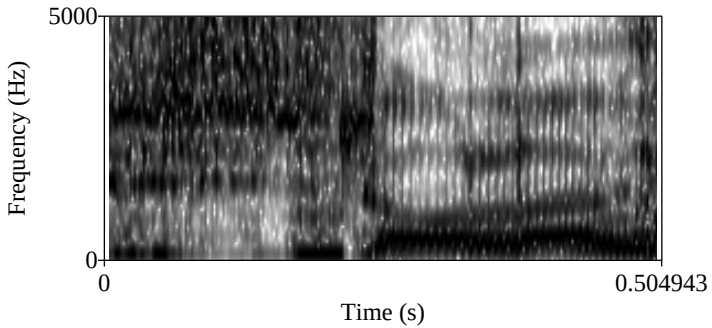


- Midden: 0 dB/octave
- Onder: +6 dB/octave

Dynamic compression

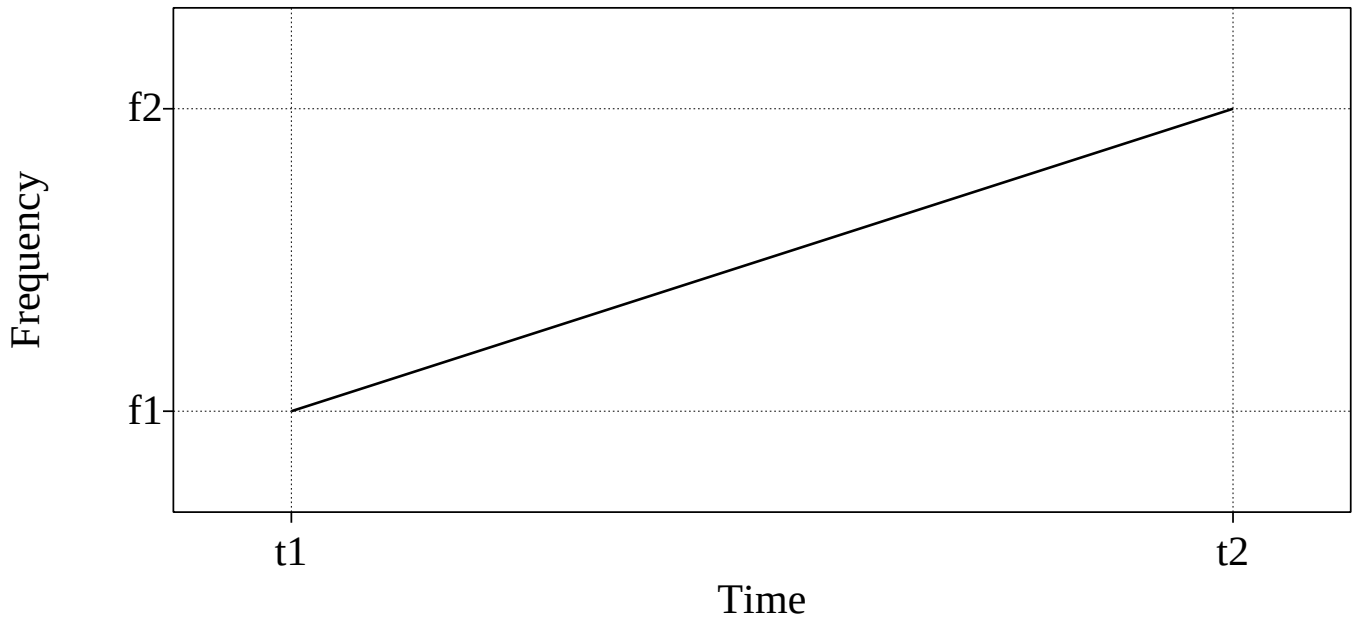






- Midden: 0 (uit)
- Onder: 1 (max)

frequentiesweep



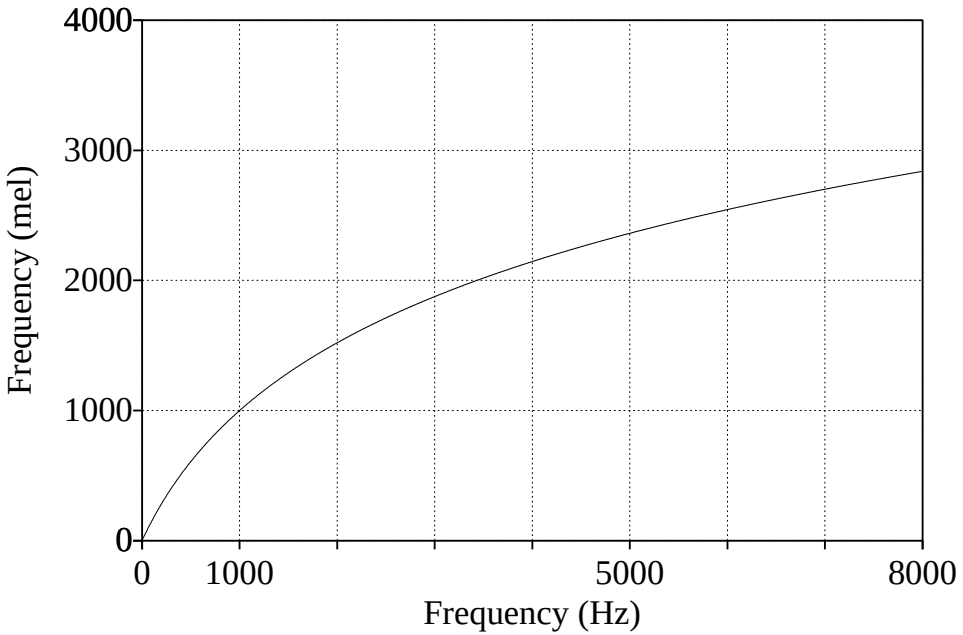
$$s(t) = \sin(2\pi((f_2 - f_1)/(t_2 - t_1)(t^2/2 - t_1 t) + f_1 t))$$

Stel `t1=0,t2=2,f1=500,f2=1500`

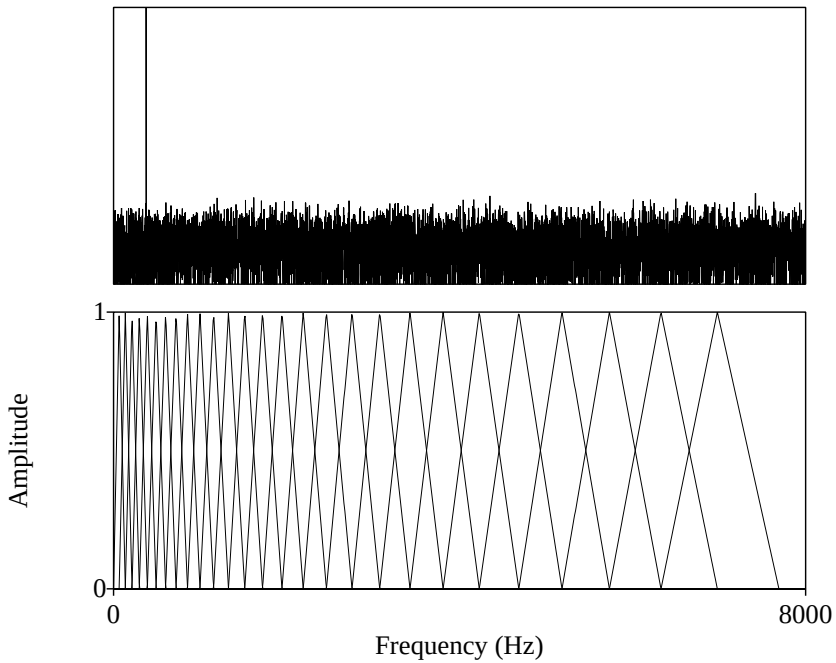
MFCC: Mel Frequency Cepstral Coefficients

Meest gebruikte representatie voor automatische spraakherkenning.

$$\text{mel}(f) = 2595 \log(1 + f/700)$$



MFCC



Berekening MFCC's per frame

1. Maak Spectrum
2. Sommeer energie via melfilters
3. Bepaal de cosinus-componenten (mfcc's)

Formanten

- Productie: Resonanties van de mond-keelholte
- Perceptie: Lokale spectrale pieken
- Akoestisch: Oplossing van een vergelijking
- Moeilijk te bepalen: kennis van "gewenste" waardes nodig

Formantdata in PRAAT


```
Create TableOfReal (van Nierop 1973)...  no
To TableOfReal (means by row labels)...  no
Draw scatter plot...
```

(& ook Pols 1973)