DSP4Linguists

Leonardo Araujo

July 30, 2021

URL

Figure 1: https://raw.githubusercontent.com/leolca/lectures/master/dsp4linguists/main.pdf

Acoustic wave

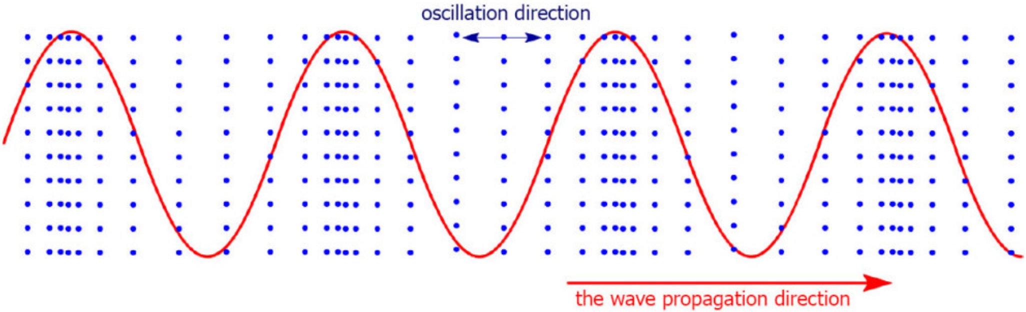


Figure 2: Acoustic wave.

Analog > Digital / Digital > Analog

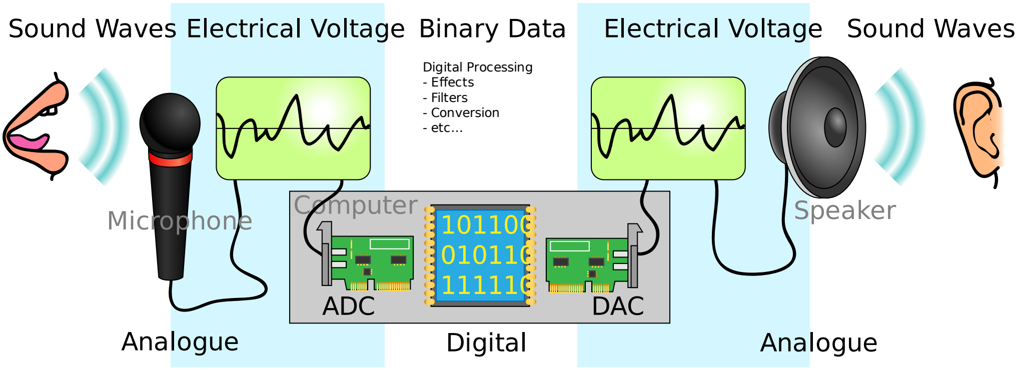


Figure 3: ADC and DAC.

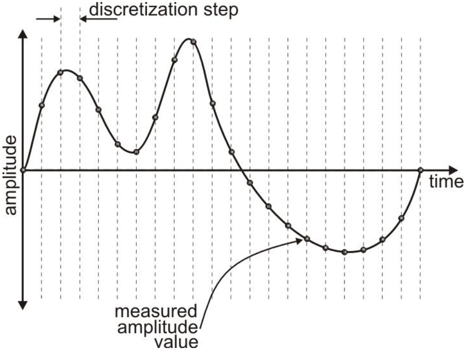


Figure 4: Sampling.

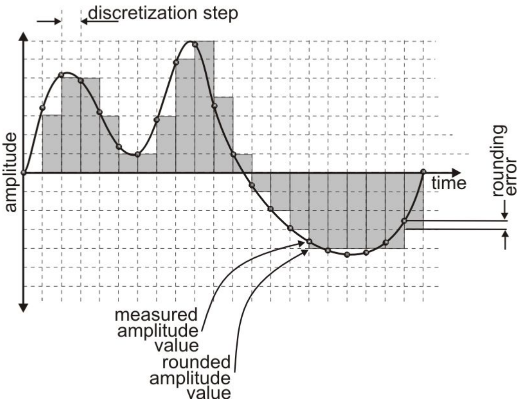


Figure 5: Quantization.

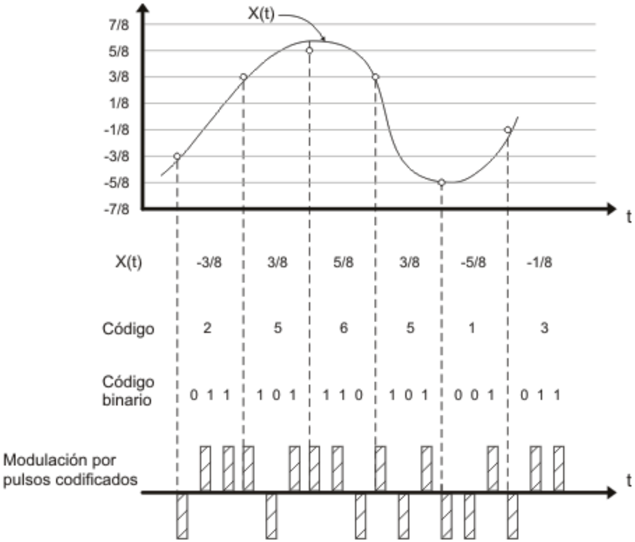


Figure 6: Coded pulses.

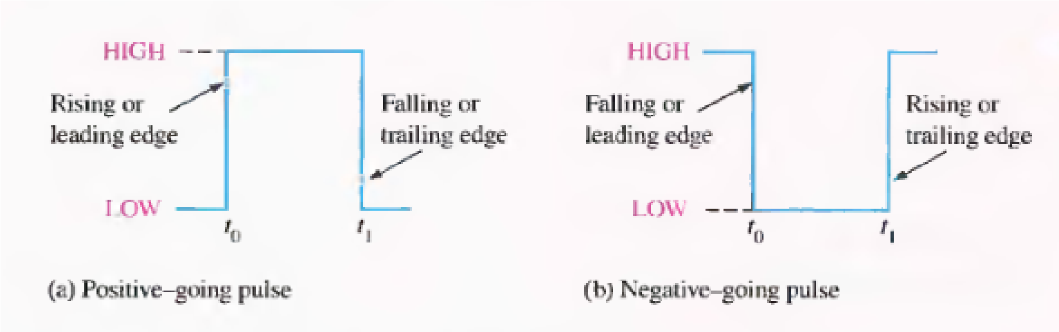


Figure 7: Digital pulse.

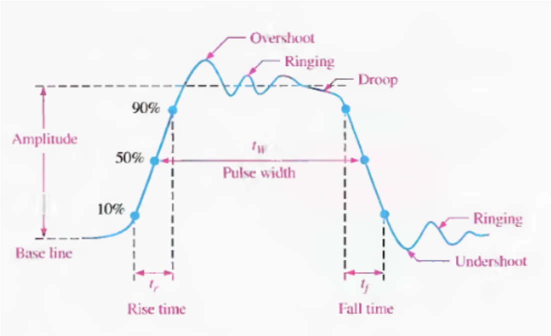


Figure 8: Non-ideal Pulse.

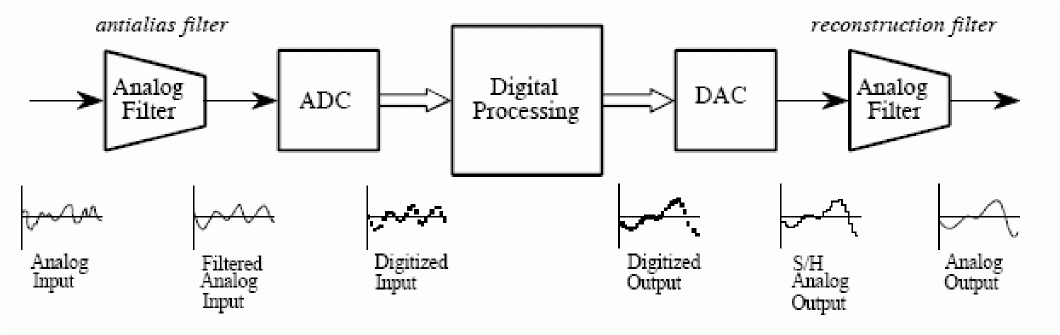


Figure 9: ADC and DAC

Quantizer

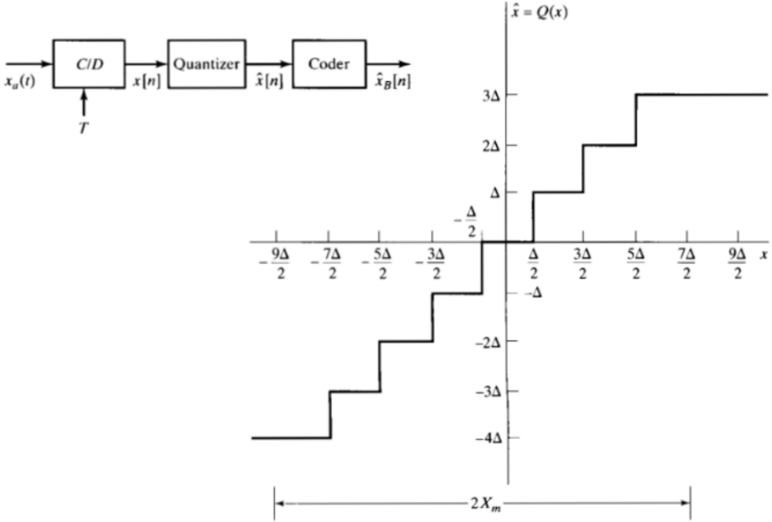


Figure 10: 3 bits uniform quantizer

Quantization examples

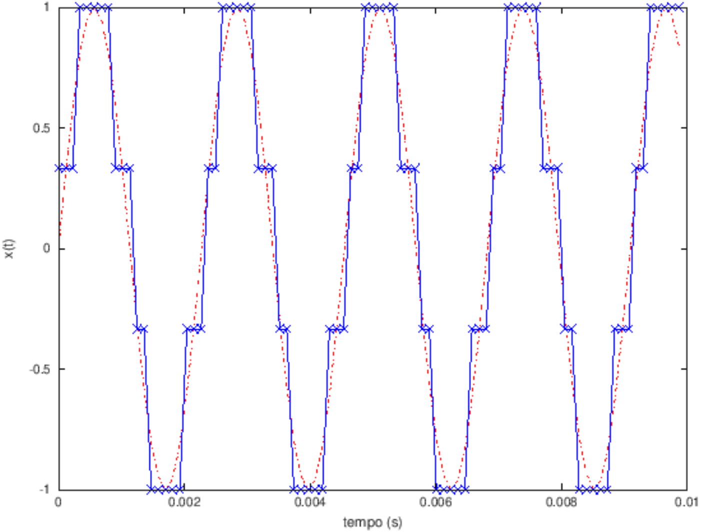


Figure 11: 440 Hz sin wave.

Clipping

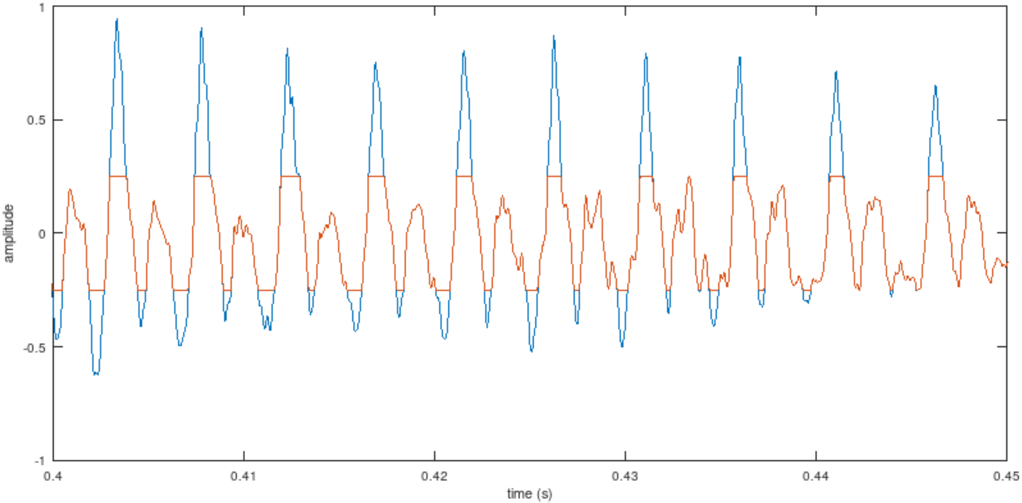
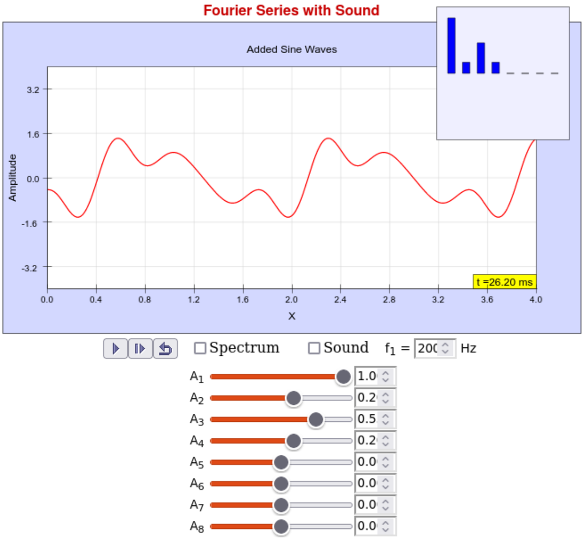


Figure 12: Signal clipping

Fourier Series

Simulation



Signals and Systems

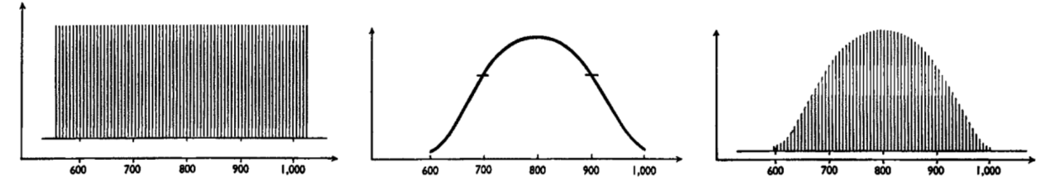


Fig. 5.3. The spectrum of a sound consisting of a large number of tones with the same amplitude.

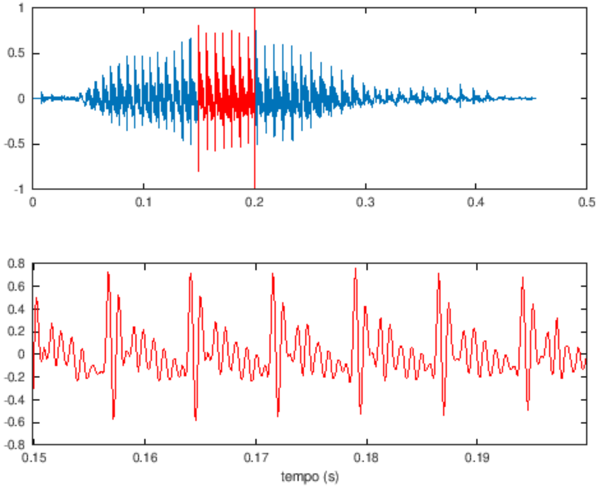
Fig. 5.4. A curve specifying a resonator.

Fig. 5.5. The output of the resonator in fig. 5.4 when the input shown in fig. 5.3 is applied to it.

Figure 14: Resonator

Vowel

open_front_unrounded.mp3



Pitch

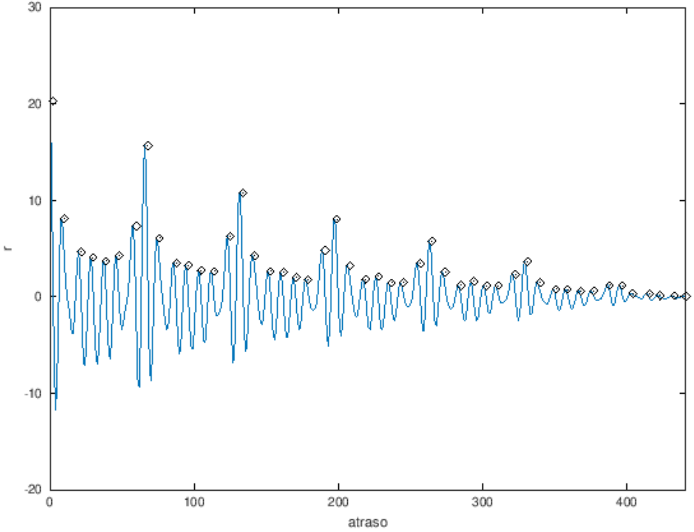


Figure 16: Autocorrelation

LPC model

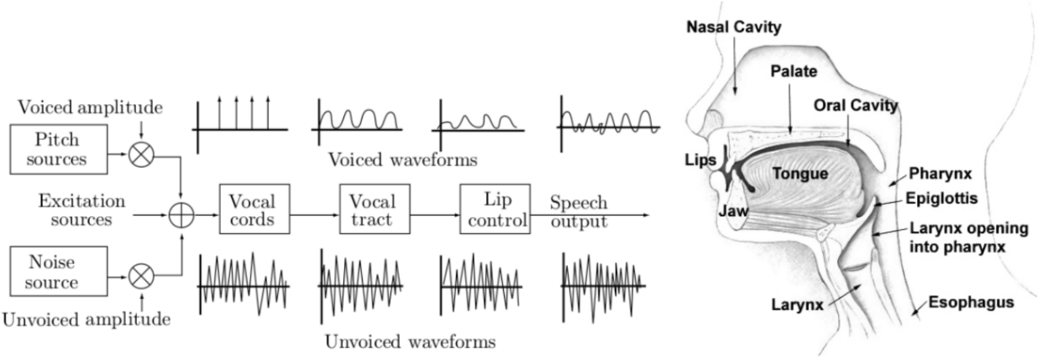


Figure 17: LPC model and vocal tract

Vocal apparatus

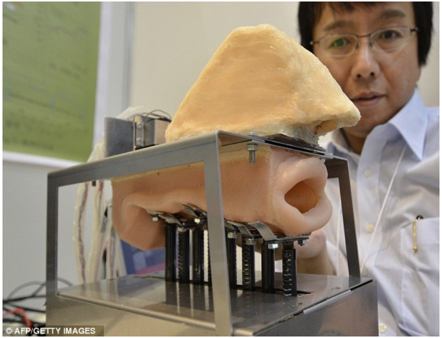


Figure 18: Hideyuki Sawada's KTR voice robot
https://www.youtube.com/watch?v=qobhDJ_vEOc

vocal cord stroboscopy examination

History of Speech Synthesis

four people sing Kyrie eleison during laryngoscopy

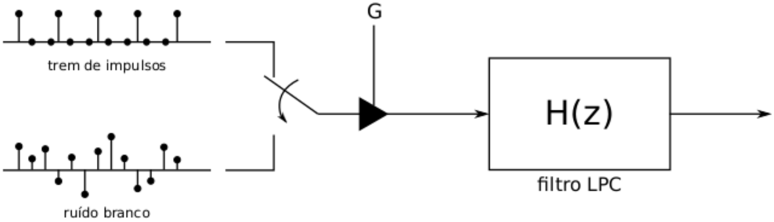


Figure 19: LPC model

synthesized example

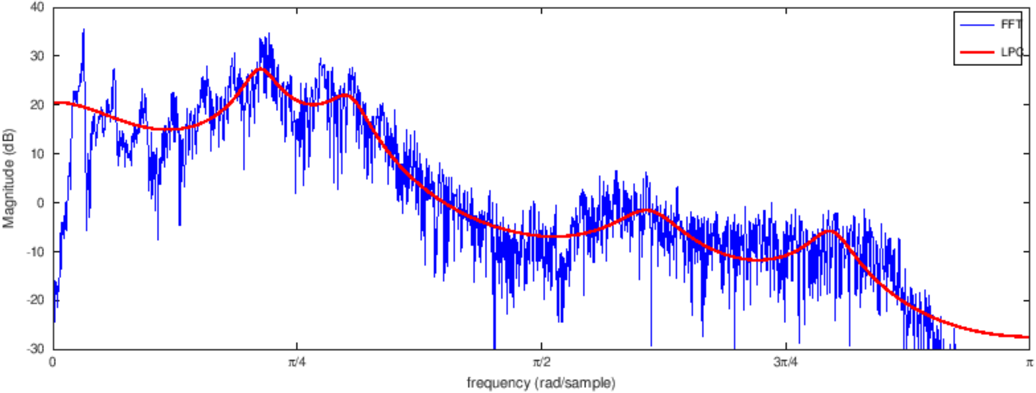


Figure 20: Spectrum and LPC.

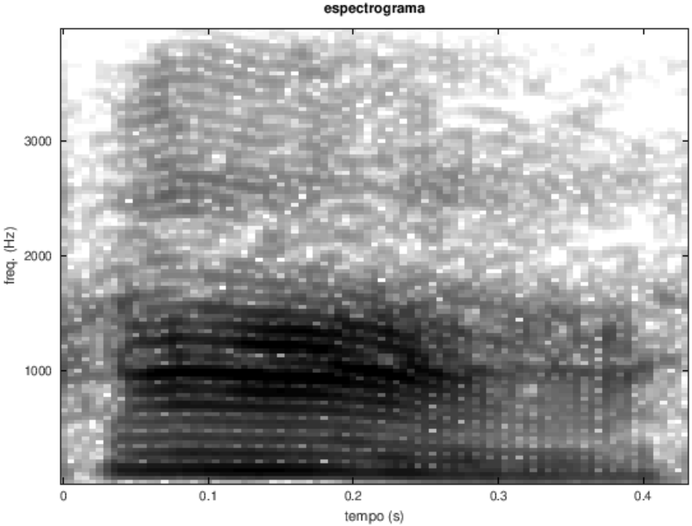


Figure 21: Spectrogram.

Spectrogram

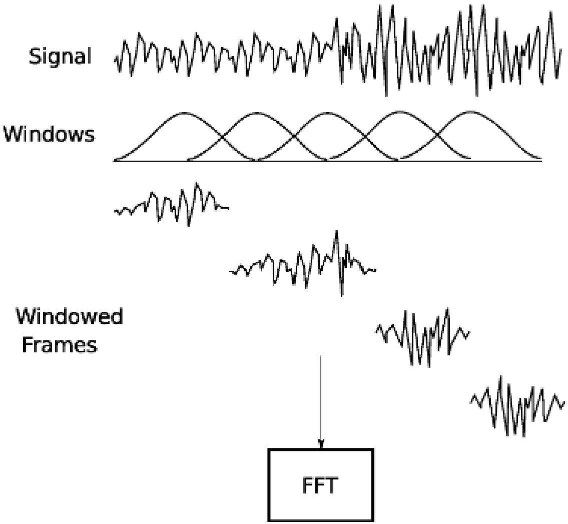


Figure 22: Schematics

Time vs Frequency

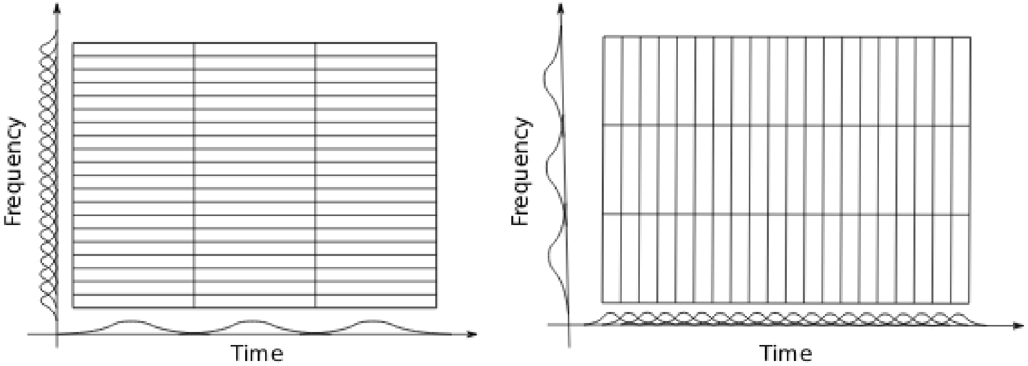


Figure 23: Uncertainty principle

The Uncertainty Principle

Figure 24: Heisenberg's uncertainty principle

Spectrogram

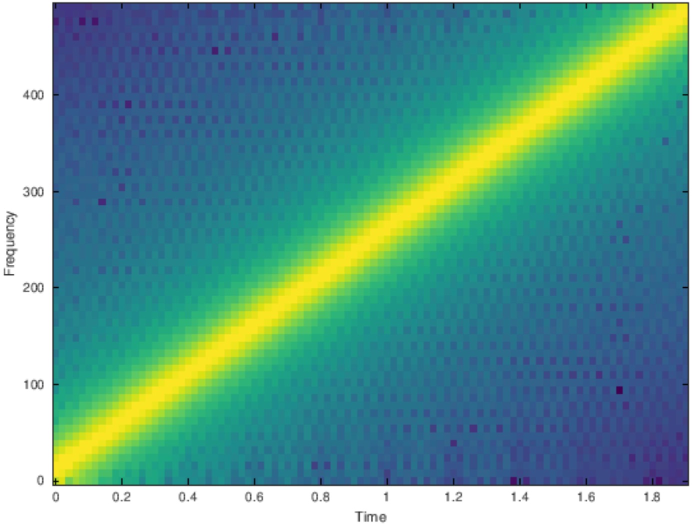


Figure 25: Chirp example, from 0 to 500Hz.

Shepard Tone

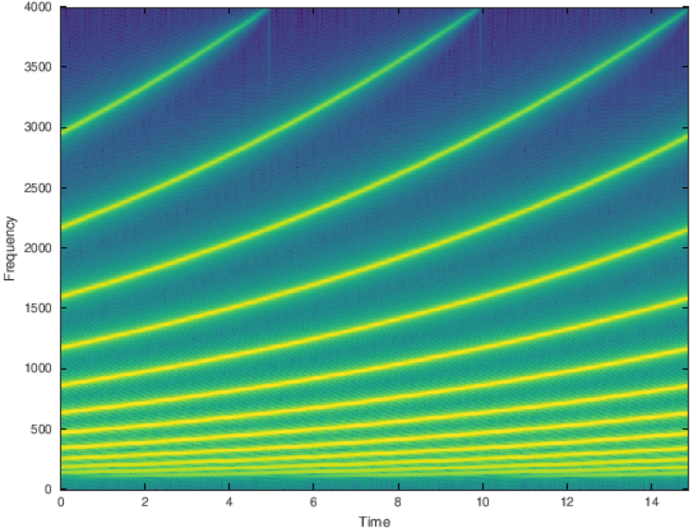


Figure 26: Shepard Tone

Downsampling

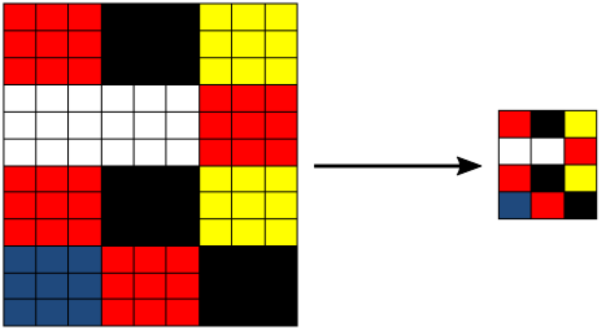


Figure 27: Downsampling example

Downsample / Decimate

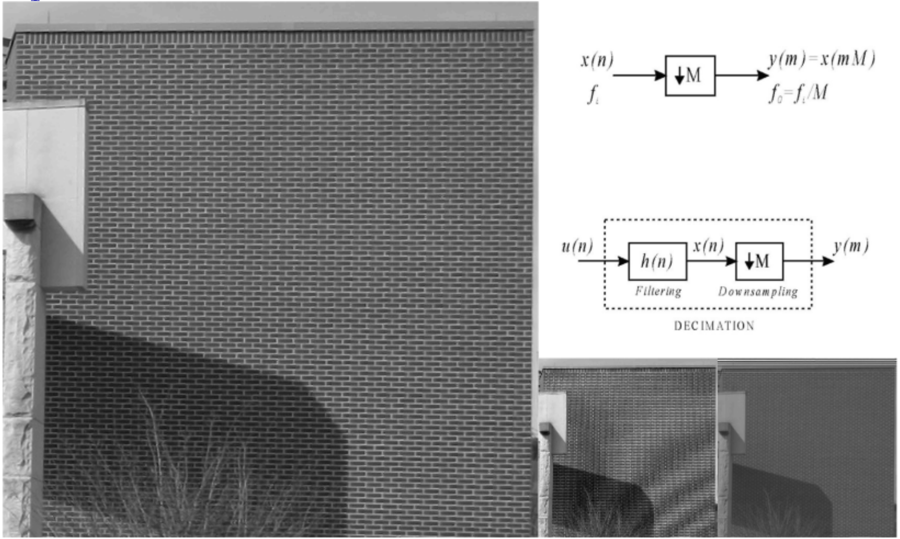


Figure 28: Downsample and Decimate

Downsample / Decimate (audio example)

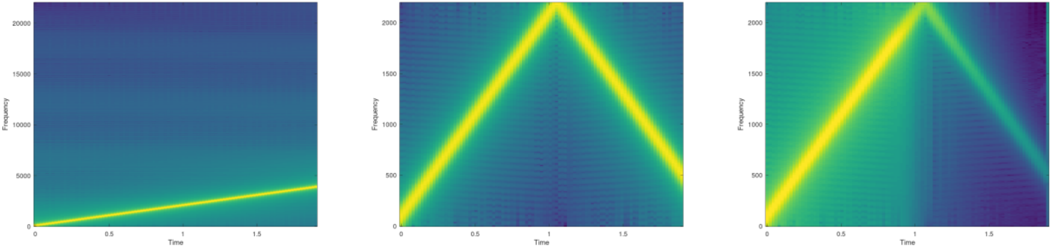


Figure 29: Downsample and Decimate

Hearing

- Intensity
- Pitch
- Duration
- Quality/timber

Pitch metamery

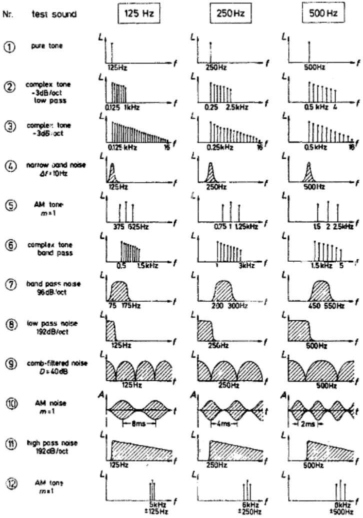


Fig. 1. Schematic representation of test sounds employed.

Figure 30: Fastl, H. & Stoll, G. Scaling of pitch strength, Hearing Research (1979): 293-301

Pitch JND

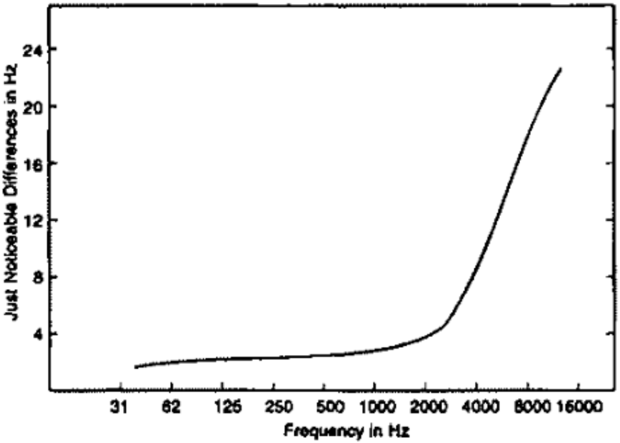


Fig. 6.3. A graph showing how much the frequency or a tone has to be altered in order to produce a change in pitch.

Figure 31: Elements of acoustic phonetics, Peter Ladefoged (1996)

Pitch scales

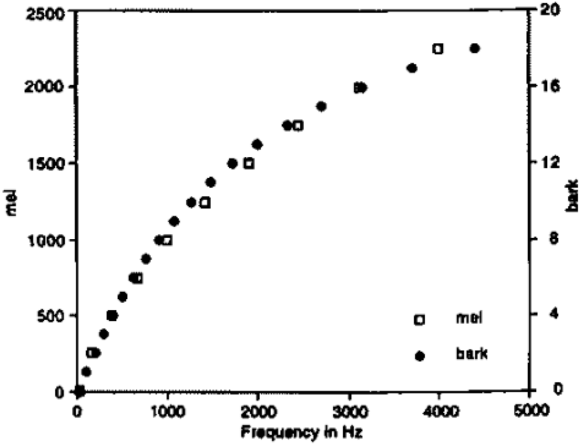


Fig. 6.4. Graph showing the relation between frequencies in Hz and the corresponding values on two different pitch scales, mel and bark.

Figure 32: Elements of acoustic phonetics, Peter Ladefoged (1996)

Duration

- Staat /ʃtaːt/ - country; state
- Stadt /ʃtat/ - city; town

Hearing range

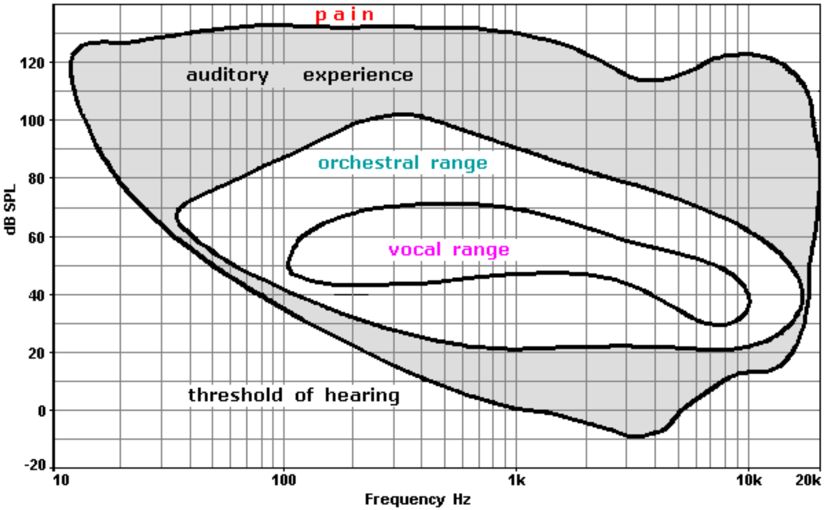


Figure 33: Hearing range

Fletcher-Munson curves

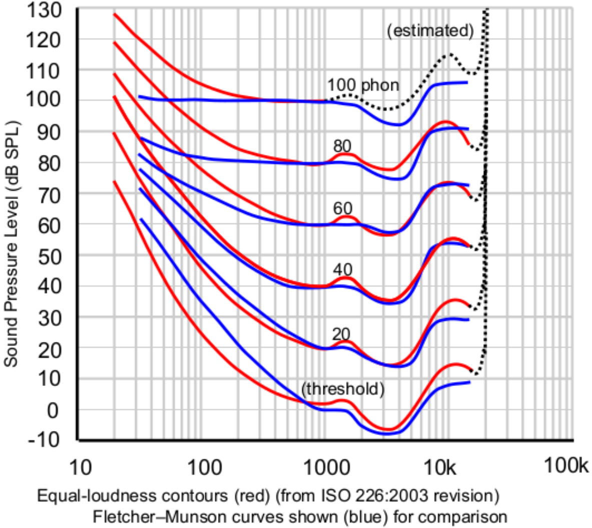


Figure 34: Equal-loudness contour

Masking

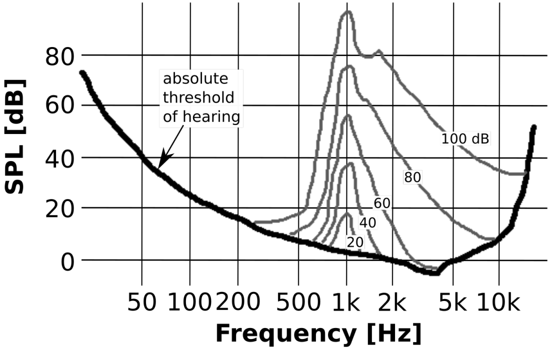


Figure 35: Frequency masking

Masking

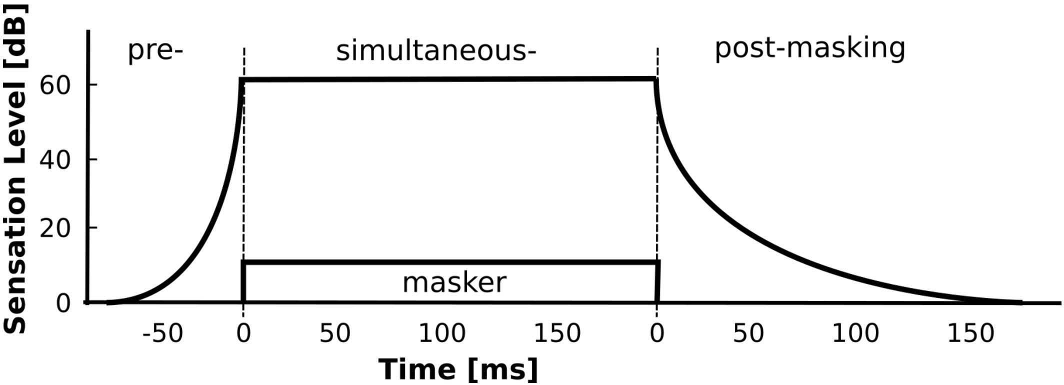


Figure 36: Time masking

Musical instruments

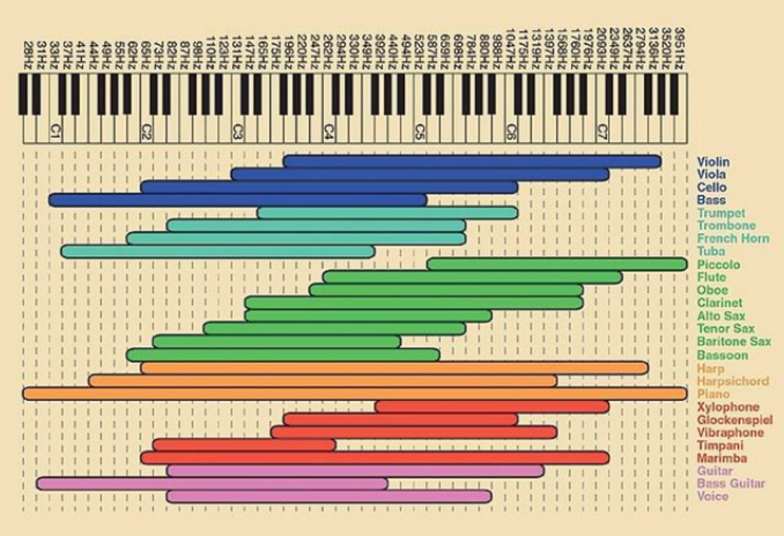


Figure 37: Frequency range of musical instruments

Ear

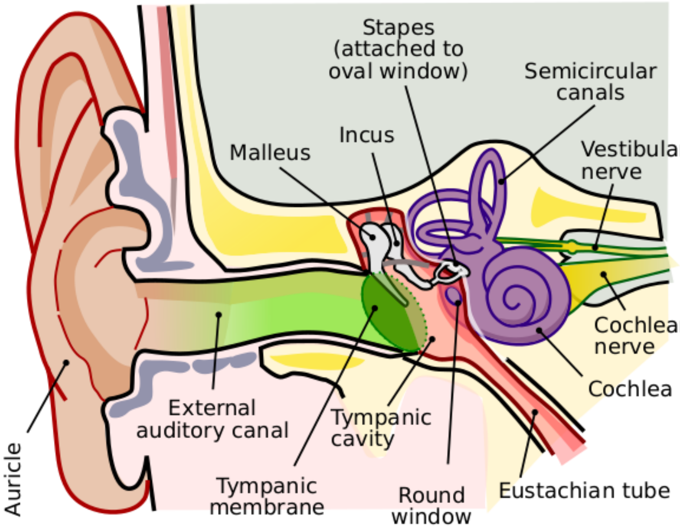


Figure 38: Ear

Middle ear

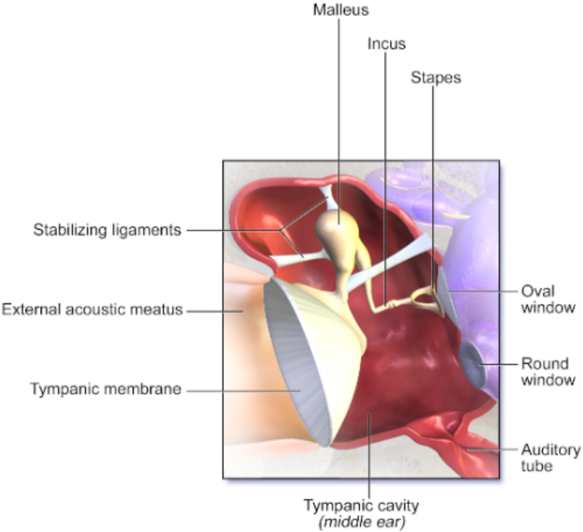


Figure 39: Middle ear

Cochlea

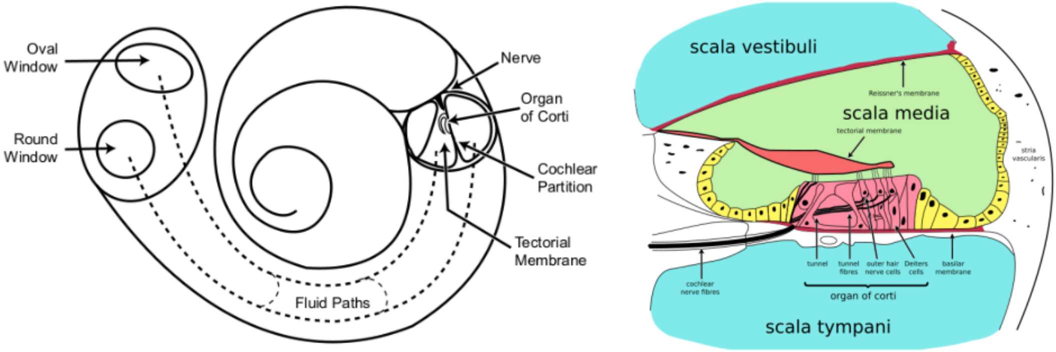


Figure 40: Cochlea and organ of Corti

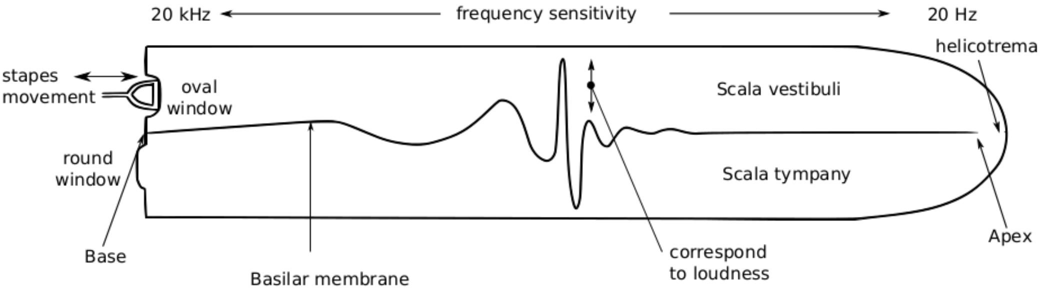


Figure 41: Travelling wave

Hair cells

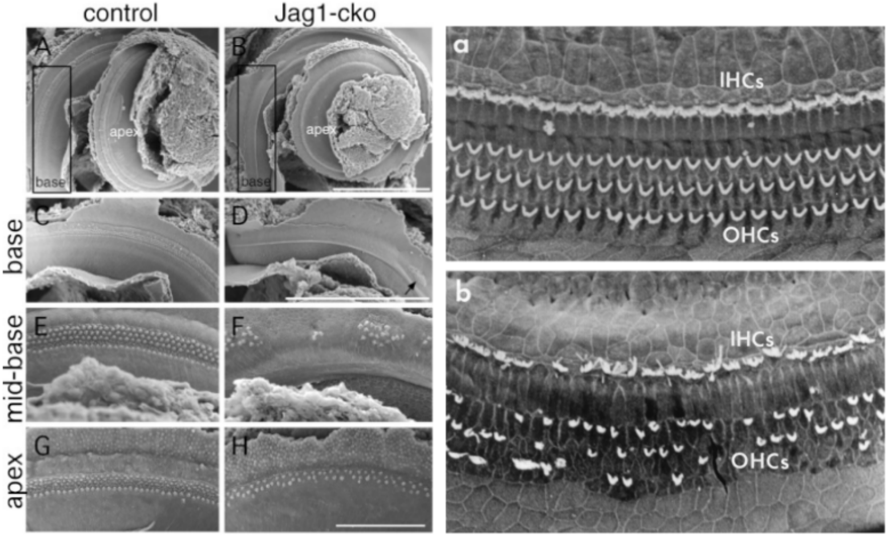


Figure 42: Inner (IHC) and outer hair cells (OHC)

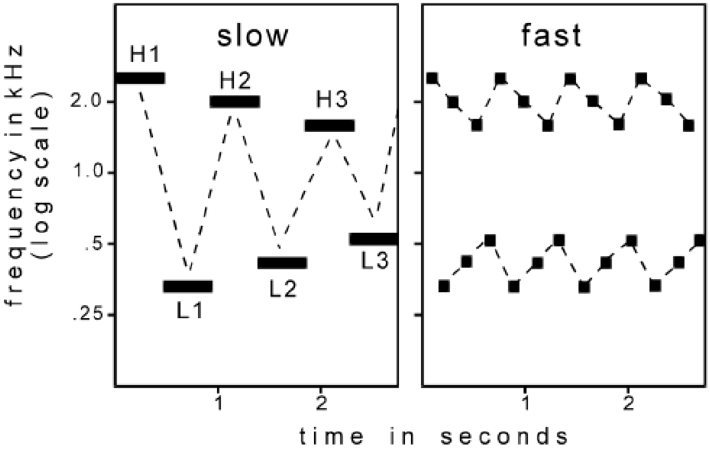


Figure 43: Stream segregation in a cycle of six tones

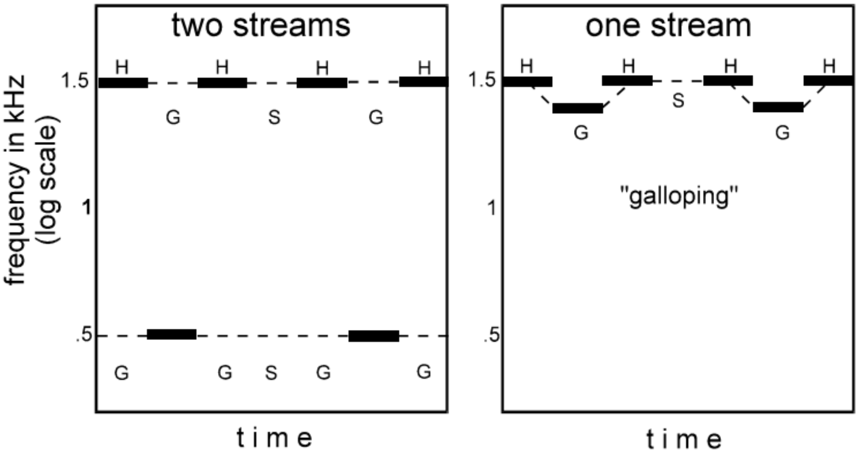


Figure 44: Loss of rhythmic information as a result of stream segregation

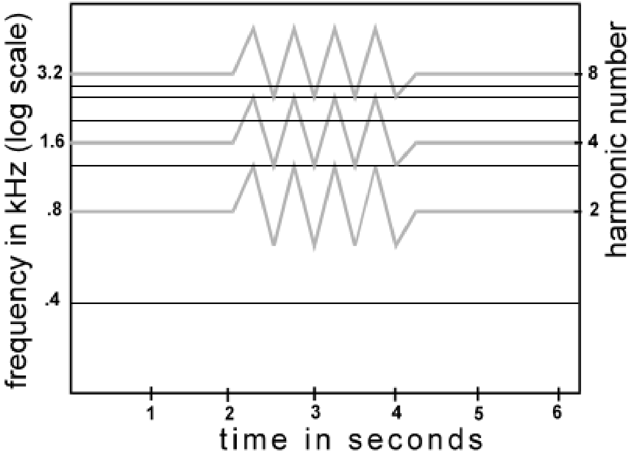


Figure 45: Fusion by common frequency change

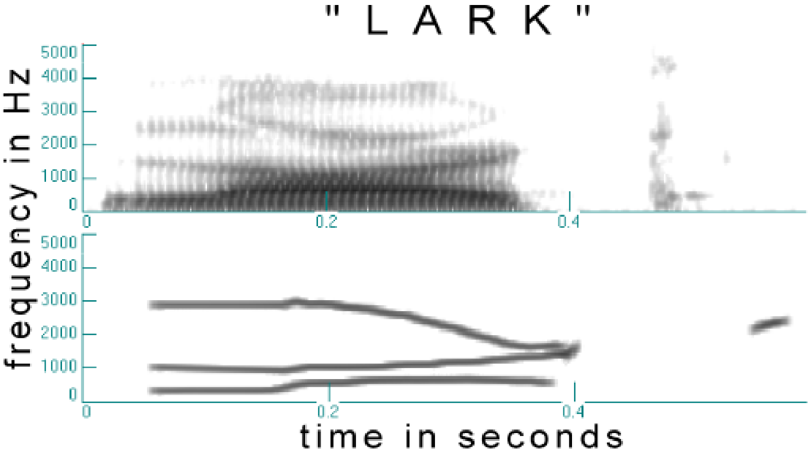


Figure 46: Sine-wave speech

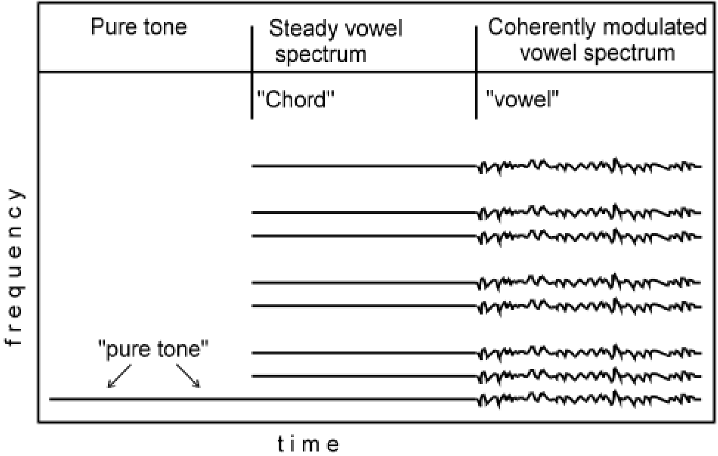


Figure 47: Role of frequency micro-modulation in voice perception

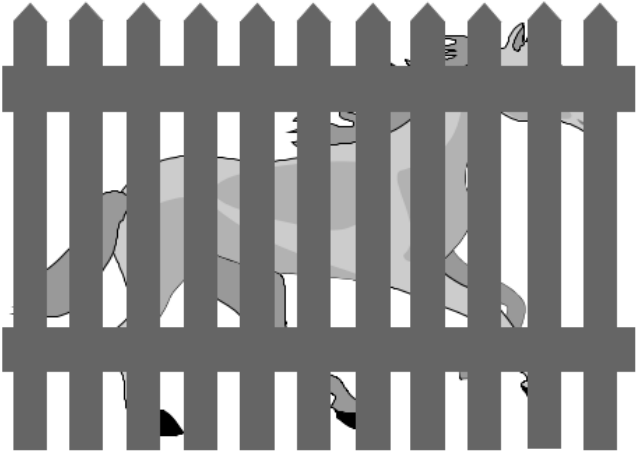


Figure 48: The picket-fence effect with speech

Audio file formats

Table 1 - Overview of selected Audio Codecs

Name	WAV	FLAC	MP3	Vorbis	AAC	Speex	Opus	WMA
Released	1991	2001	1993	2000	1997	2003	2012	1999
Compression	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Loss-less	–	Yes	No	No	No	No	No	No
Bit-rate (kbit/s)	1,411.2	935	16–320	48–500	16–320	2–24	8–128	32–448
Encoder	–	flac	lame	oggenc	ffmpeg	speexenc	opusenc	ffmpeg
Decoder	–	ffmpeg	lame	oggdec	ffmpeg	speexdec	opusdec	ffmpeg

Figure 49: I Siegert, AF Lotz, LL Duong, A Wendemuth, Measuring the impact of audio compression on the spectral quality of speech data, 2016

Compression ratio

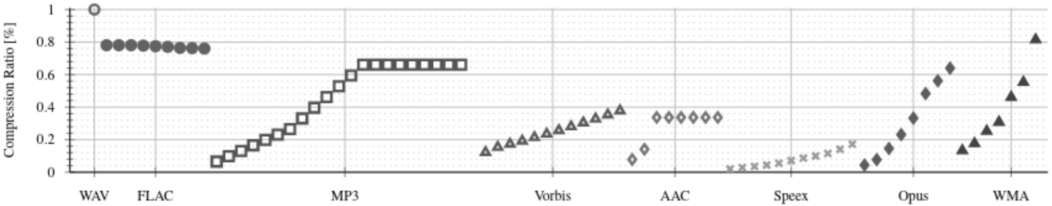


Figure 1 - Achieved average compression ratio for each codec and bit-rate. The bit-rate is increasing from left to right, see Table 2.

Figure 50: I Siegert, AF Lotz, LL Duong, A Wendemuth, Measuring the impact of audio compression on the spectral quality of speech data, 2016

Compresstion error

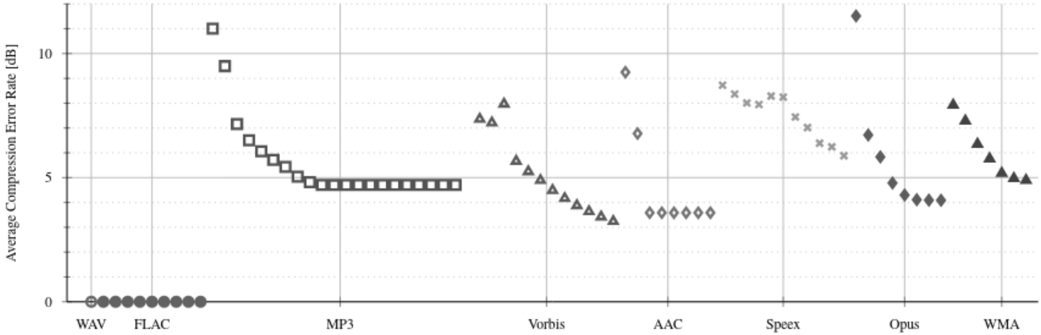


Figure 3 - Average compression error rate for each codec and bit-rate. The bit-rate is increasing from left to right, see Table 2.

Figure 51: I Siegert, AF Lotz, LL Duong, A Wendemuth, Measuring the impact of audio compression on the spectral quality of speech data, 2016

Compression ratio vs compression error

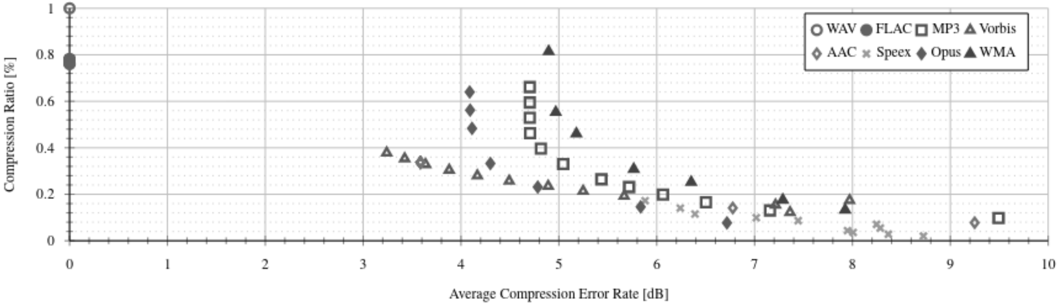


Figure 4 - Average compression ratio over average compression error rate for each codec and bit-rate.

Figure 52: I Siegert, AF Lotz, LL Duong, A Wendemuth, Measuring the impact of audio compression on the spectral quality of speech data, 2016

Conclusion

“we recommend to use FLAC for all cases where the accuracy matters. In cases where a slight error is acceptable, we recommend Vorbis at 500 kbit/s” (Siegert et al 2016)

Praat

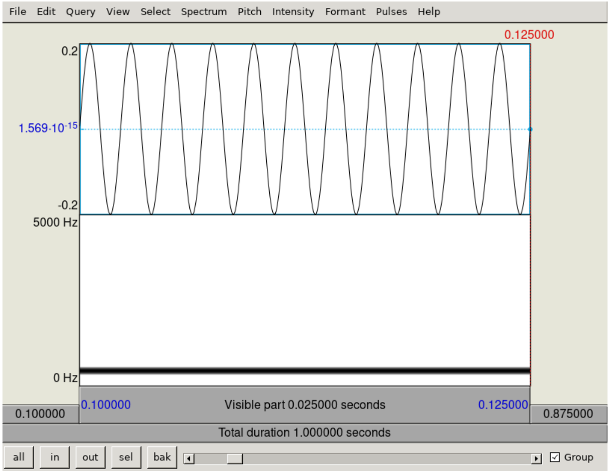


Figure 53: 440Hz sin wave

Praat

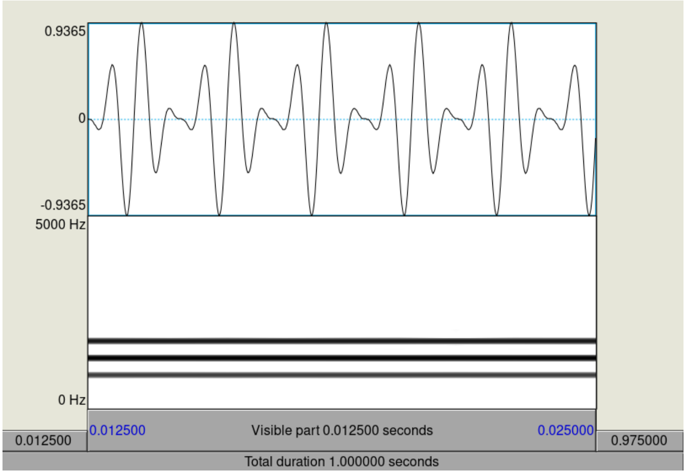


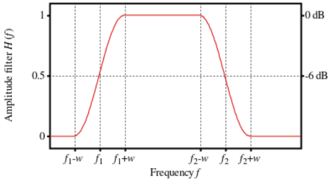
Figure 54: 1/4 sin(2 pi 880x) + 1/2 sin(2 pi 1320x) + 1/4 sin(2 pi 1760x)

Praat

Spectrum: Filter (pass Hann band)...

A command to modify every selected **Spectrum** object.

The complex values in the **Spectrum** are multiplied by real-valued sine shapes and straight lines, according to the following figure:



Settings

From frequency (Hz) (standard value: 500 Hz)

the lower edge of the pass band (f_1 in the figure). The value zero is special: the filter then acts as a low-pass filter.

To frequency (Hz) (standard value: 1000 Hz)

the upper edge of the pass band (f_2 in the figure). The value zero is special: the filter then acts as a high-pass filter.

Smoothing (Hz) (standard value: 100 Hz)

the width of the region between pass and stop (w in the figure).

Usage

Because of its symmetric Hann-like shape, the filter is especially useful for decomposing the Spectrum into consecutive bands. For instance, we can decompose the spectrum into the bands 0-500 Hz, 500-1000 Hz, 1000-2000 Hz, and 2000-"0" Hz:

Figure 55: Filter pass Hann band

Praat

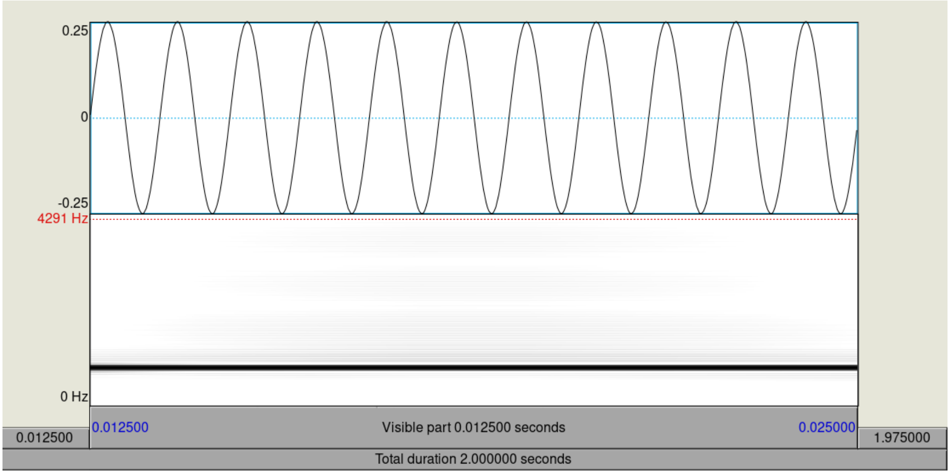


Figure 56: f1=780, f2=980, w=100

Praat

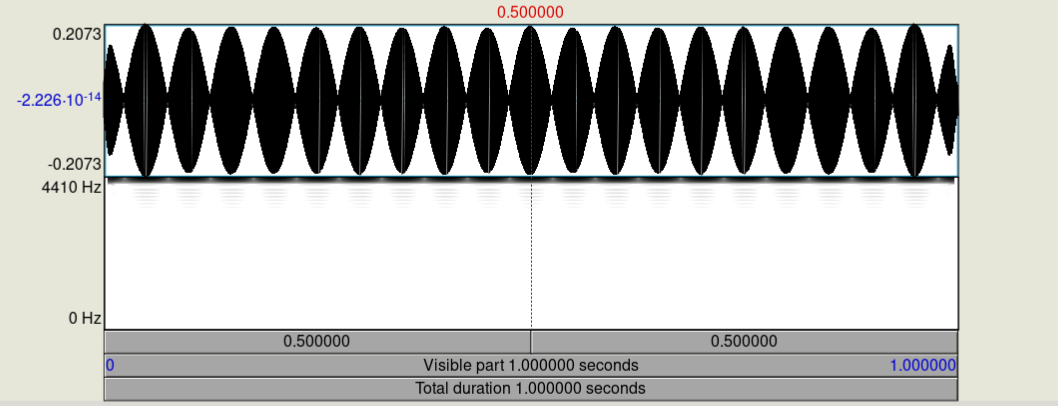


Figure 57: 4400Hz sin wave

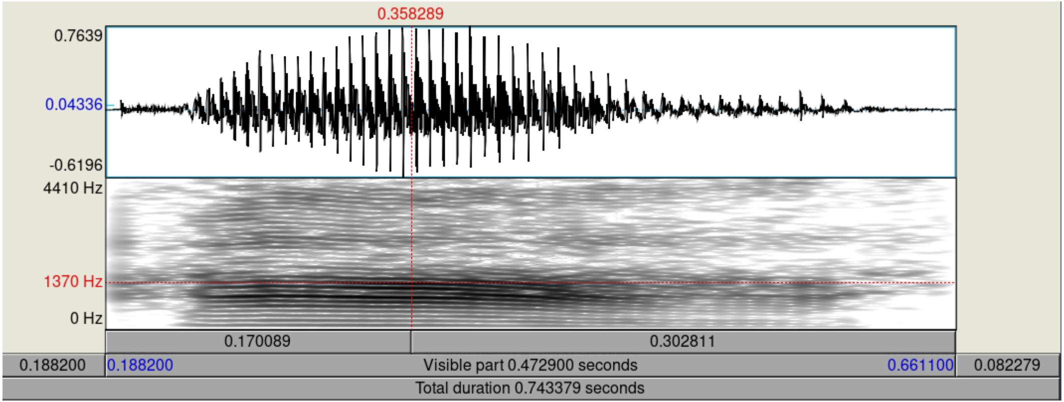

0.358289
0.7639
0.04336
-0.6196
4410 Hz
1370 Hz
0 Hz
0.170089
0.302811
0.188200
0.188200
Visible part 0.472900 seconds
0.661100
0.082279
Total duration 0.743379 seconds

Figure 58: ah

Praat

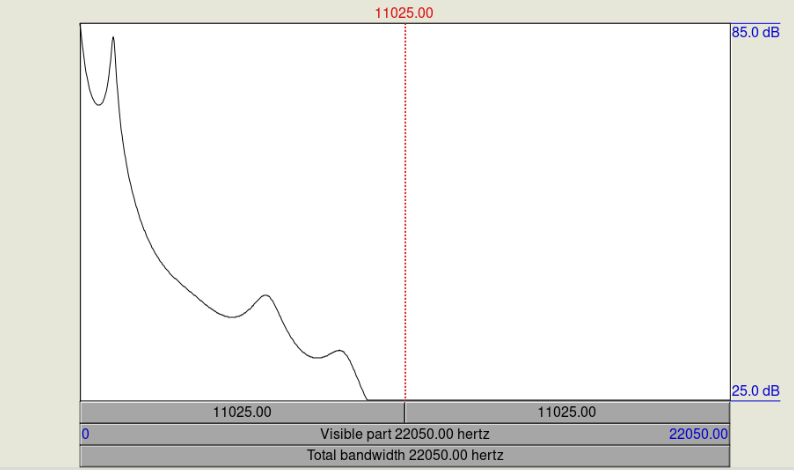


Figure 59: ah

Praat

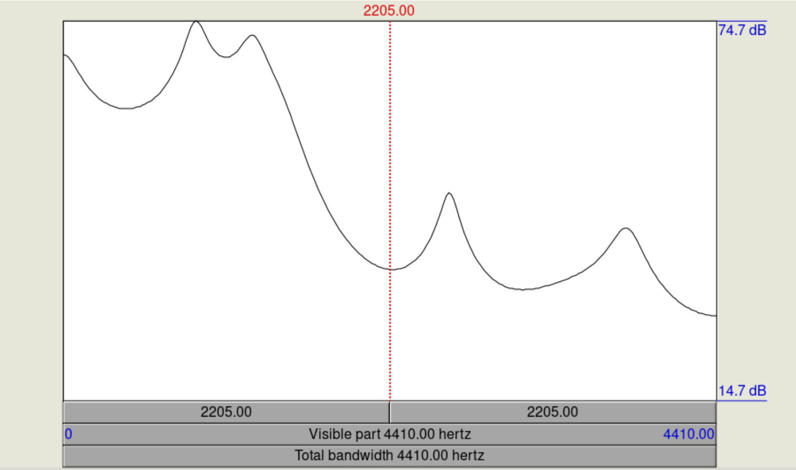


Figure 60: ah

THE END