Speech information systems

1. Basic speech acoustics

Basic concepts

Language: the most important tool for human thought and communication

Speech: the primary form of language expression

Multimodality: acoustic – visual

haptic

feeling of heat

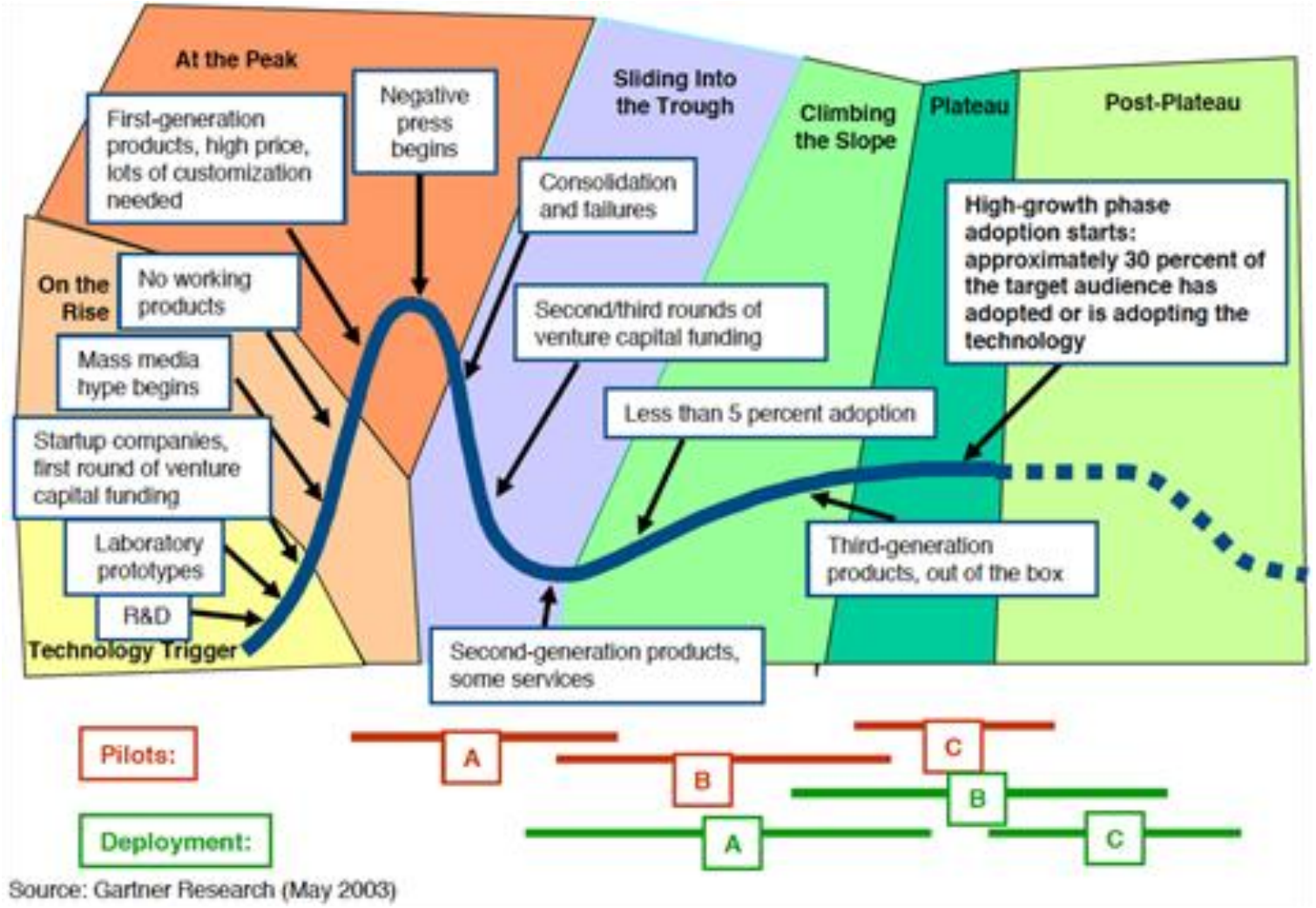
smell

balance

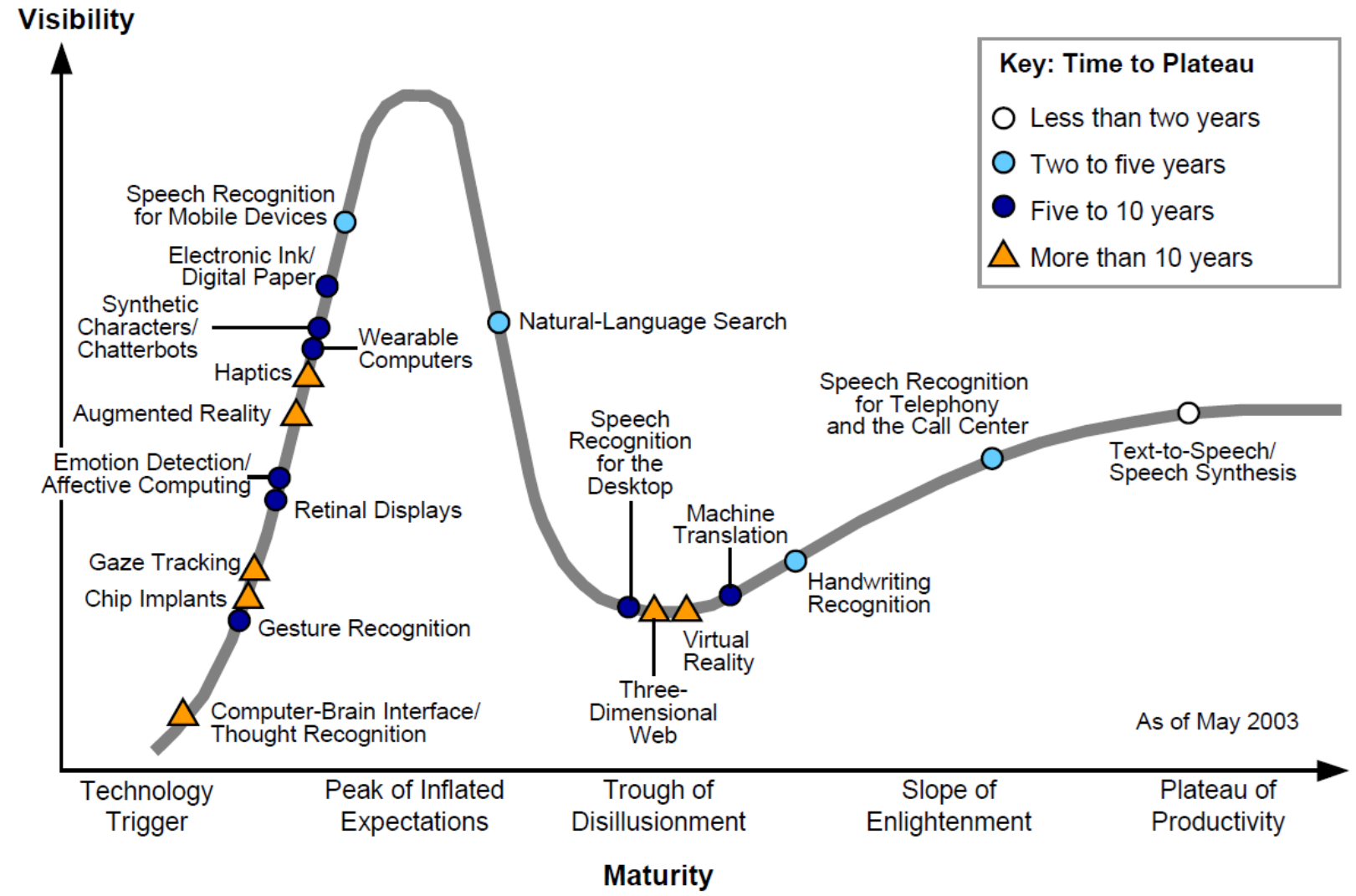
……..

Gartner Hype Cycle

- What is this, in general?



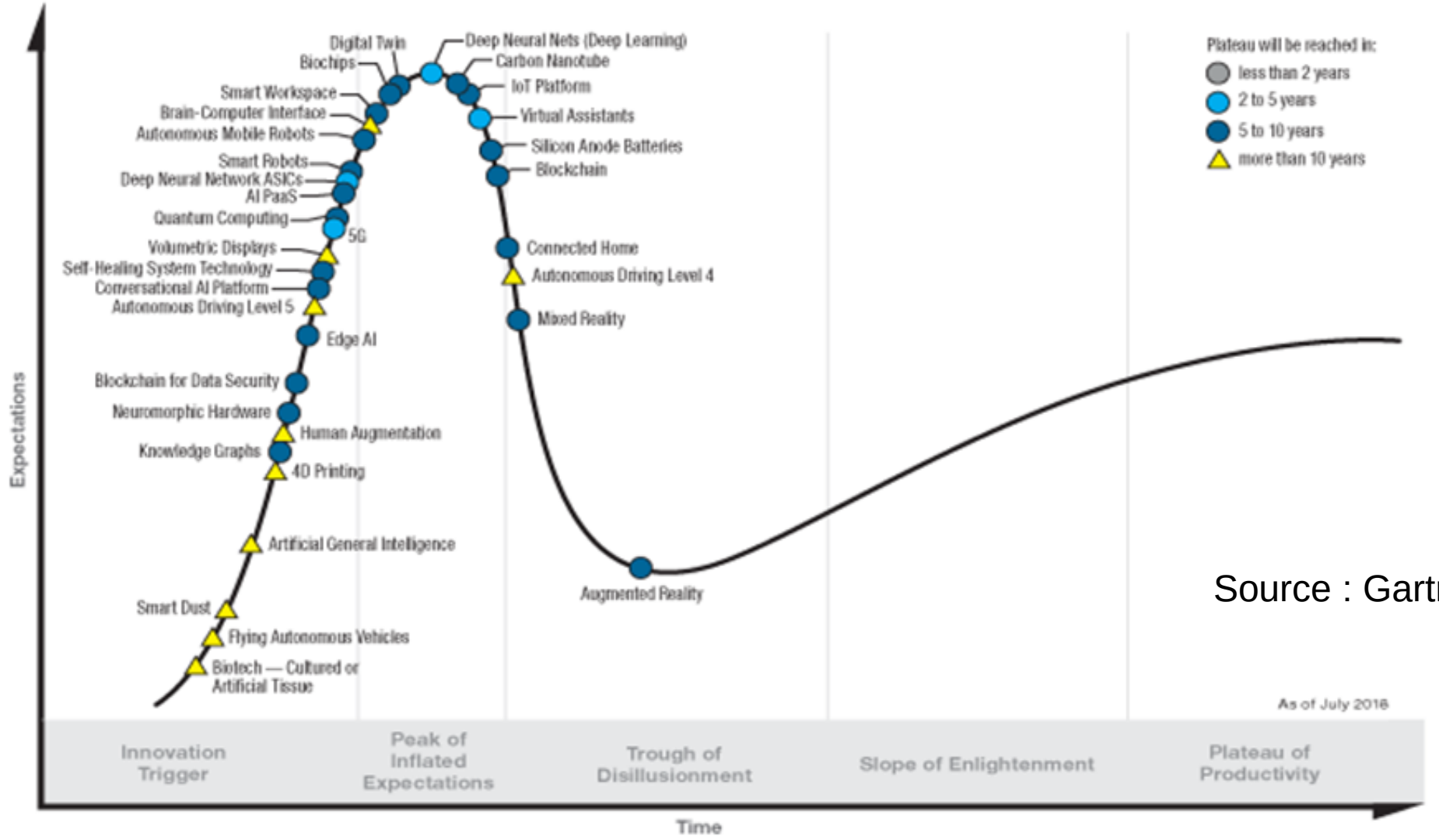
Old Hype Cycle



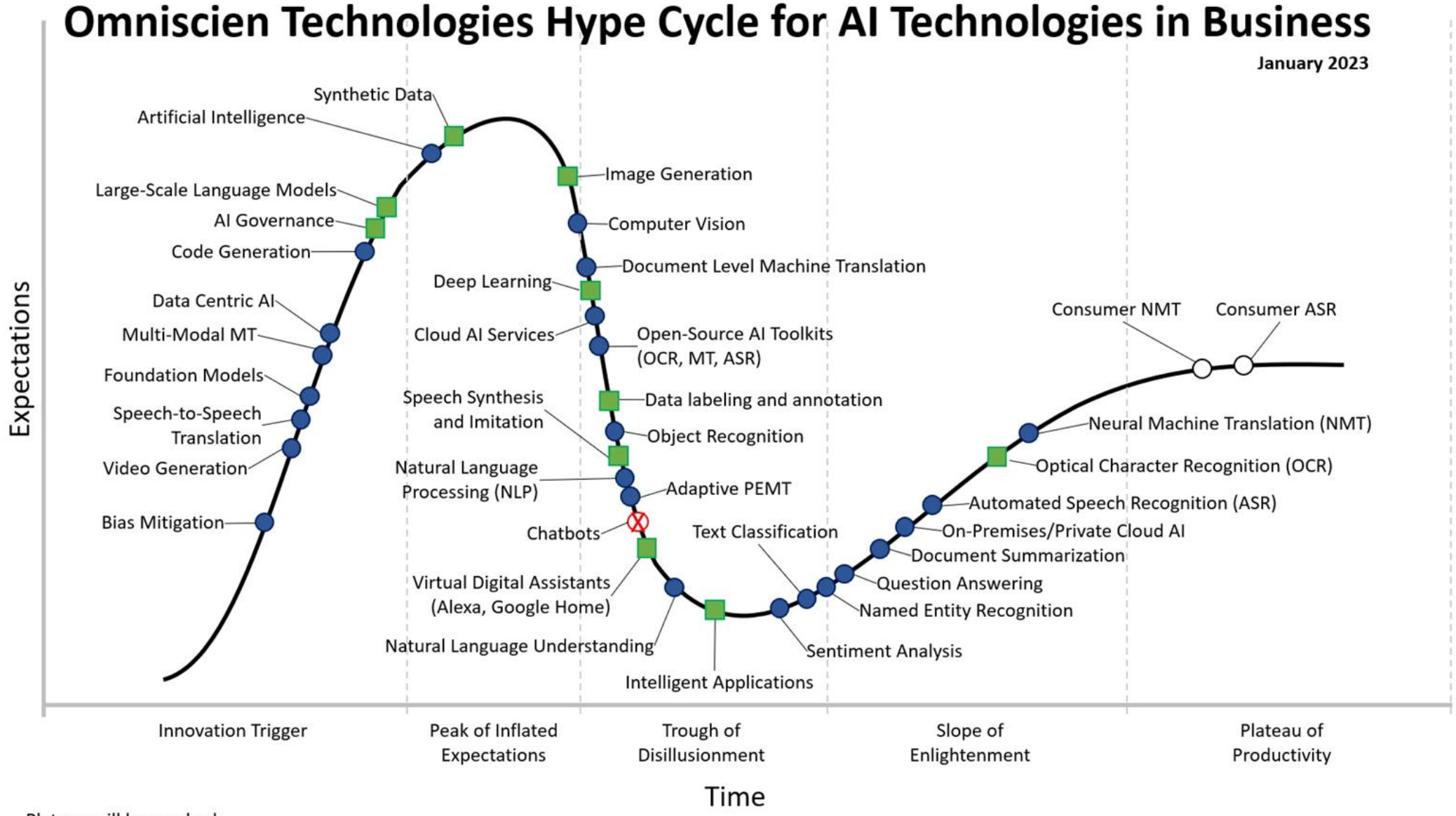
Source: Gartner Research (May 2003)

Figure 1. Hype Cycle for Human-Computer Interaction, 2003

Hype Cycle for Emerging Technologies, 2018

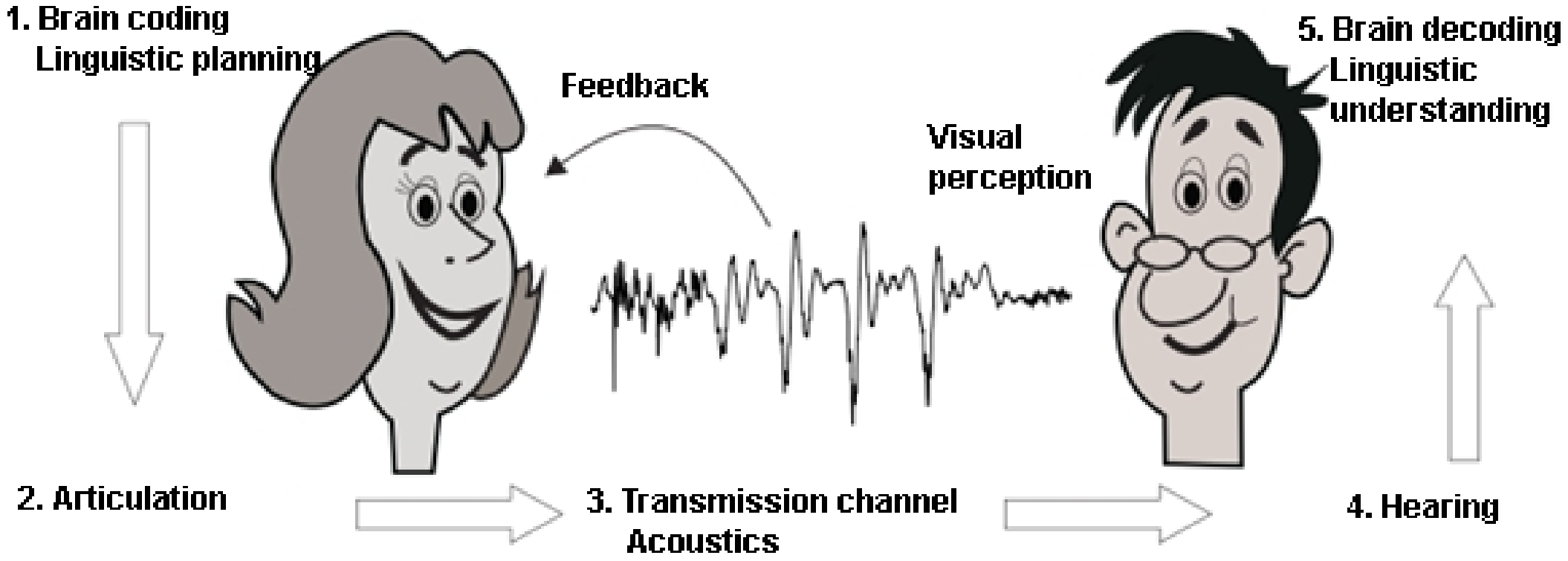


Source : Gartner August 2018

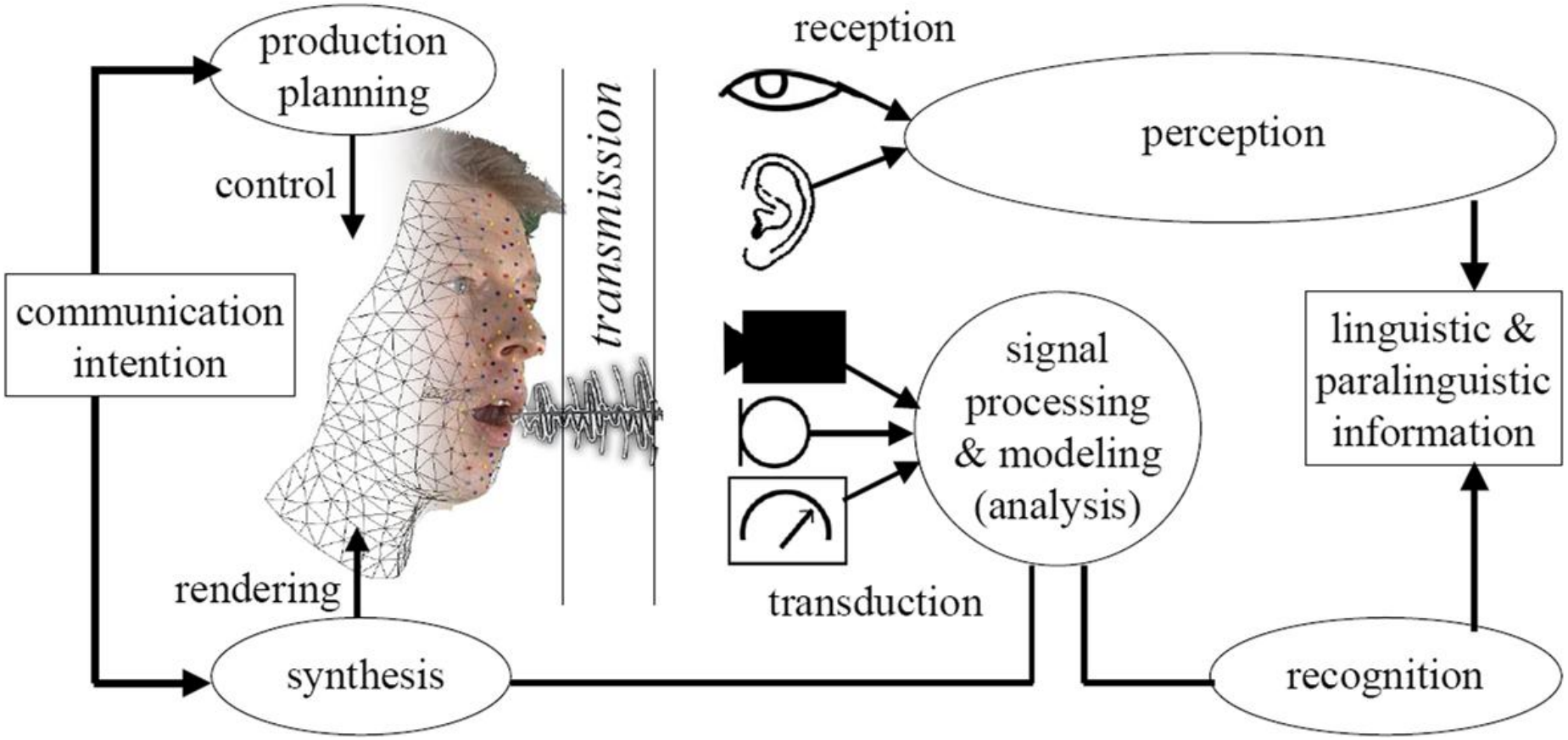


Source: https://omniscien.com/blog/hype-cycle-for-ai-technologies-in-business/

The communication model of the natural speech chain



The communication model of the natural speech chain



Source: Fagel (2007)

Speech…..

- Speech processing
  - Examination and modeling of the elements of the natural speech chain
  - <> natural language processing , NLP
  - Text <> language <> speech, chatbot <> voice bot
- Speech technology
  - Machine implementation of one or more elements of the natural speech chain
- Speech synthesis / speech generation
  - Machine generation of waveforms from symbolic input
  - Text-to-Speech (TTS) <> screen-reader
  - concept-to-speech (CTS) with additional information
- Speech recognition
  - Machine recognition of speech content
  - Speech-to-Text conversion (speech-to-text, STT, ASR)

Speech…

- Speech understanding
- Speaker recognition
  - Speaker verification
  - Speaker identification
  - Speaker change detection
  - Speaker diarization
- Speech information system
  - Information system created using speech technology elements

Milestones
(www.magyarbeszed.hu)

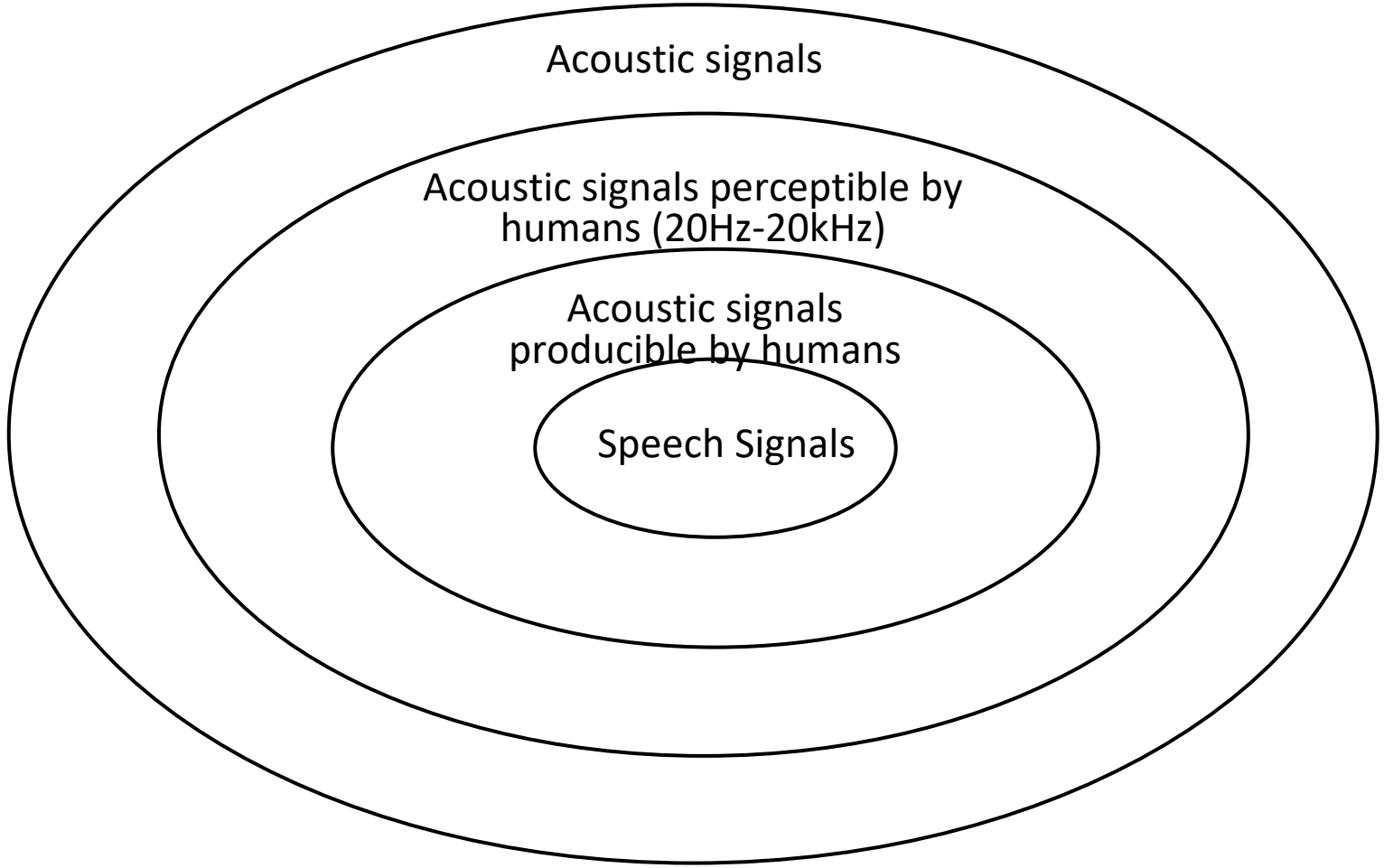
- 1791 Wolfgang von Kempelen, mechanical speech synthesis
- 1876 Alexander Graham Bell
- 1900 Helmholtz, resonators
- 1893 Tivadar Puskás, telephone exchange, telephone announcer
- 1916-19 Miklós Bánó, electromechanical speech synthesis
- 1938 Reeves, PCM (basics of digital speech transmission)
- 1939 Dudley, VODER (electric talking machine based on Helmholtz principles)
- 1947-50 György Békésy, research on the mechanism of hearing,
  PKI – 1960 Nobel Prize

Milestones

- 1950 Dennis Gabor, time-frequency resolution of the ear
- 1980 Denis Klatt , MITalk , DecTalk , first English commercial TTS
- 1982 Gábor Olaszy – Gábor Kiss, HungaroVox , first Hungarian TTS
- 1999 BME TMIT - HTE Eurospeech conference
  - BME TMIT – Westel, MAILMONDÓ, Hungarian-language, commercial email reading service
- 2004 BME TMIT – MIT Systems Kft – Westel, SMSMondó – SMSRapper , the world's first commercial smartphone SMS reader service
- 2007 BME TMIT – OGYI, Medicine Line, patient information telephone-web information system
- .....

2. Basic acoustic concepts

Wave propagation model of acoustic signals



Physical description of sound

- Speed of sound (*c*):

$$c = \frac{\lambda}{T} = \lambda f [m/s].$$

Wavelength (λ) : is the distance covered by the sound wave during one period (T)

f = 1/T, In air *c* = 344 m/s, at 20C °, 1 [atm] = 100,000 [Pa] pressure

Wavelength of a 1kHz whistle: 344m/s/1000Hz=34.4cm

- The sound pressure: $p(t) = p_0 + p_{speech}(t)$

$$p_{eff} = \sqrt{p^2(t)} = \sqrt{\frac{1}{t_2 - t_1} \int_{t_1}^{t_2} p^2(t) dt} \quad [N/m^2].$$

Physical description of sound

- Speed of sound (*c*): $$c = \frac{\lambda}{T} = \lambda f [m/s].$$

The wavelength of the sound (λ) is the distance traveled by the soundwave during one period (T).

f = 1/T, In air *c* = 344 m/s, at 20C ^{o} , 1 [atm] = 100,000 [Pa] pressure

Wavelength of a 1kHz whistle: 344m/s/1000Hz=34.4cm

- The sound pressure: $p(t) = p_0 + p_{speech}(t)$

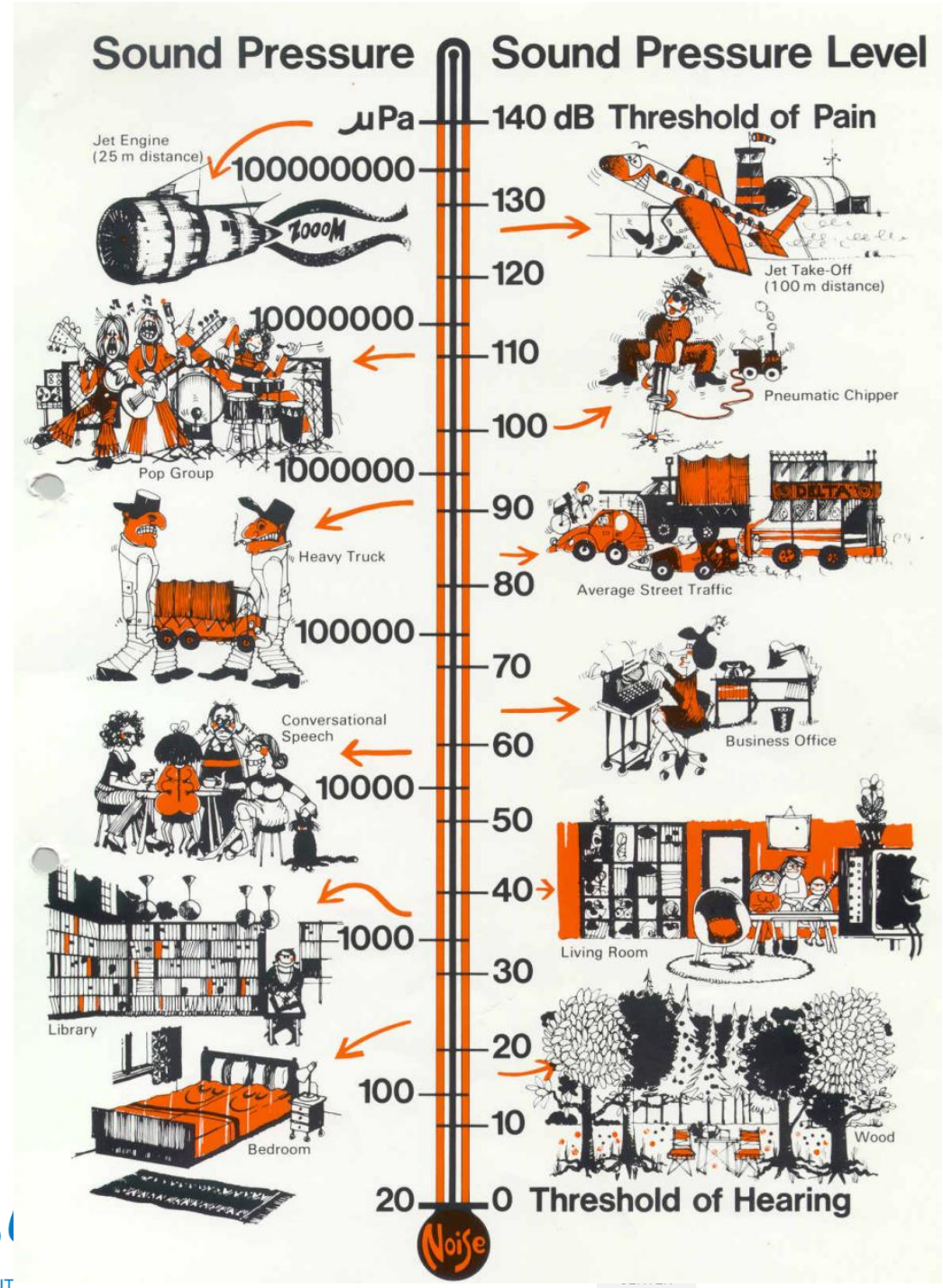
$$p_{eff} = \sqrt{p^2(t)} = \sqrt{\frac{1}{t_2 - t_1} \int_{t_1}^{t_2} p^2(t) dt} \quad [N/m^2].$$

Sound pressure, Intensity, levels

- $p = F/A = m * a / A$ [kg * m/s^2 / m^2] P= [kg / m * s^2] = [Pa]
- $I = P/A$ = [W/m^2]
- Sound pressure level
  - $L = 20 \lg p$ [Pa] / $20*10^{-6}$ [Pa] = $10 \lg I$ [W/m^2] / 10^{-12} [W/m^2] = [dB$_a$]

3. The auditory system

Dynamic range of hearing



Source:
http://personal.cityu.edu.hk/~bsapplec/sound.htm

Some characteristics of human hearing

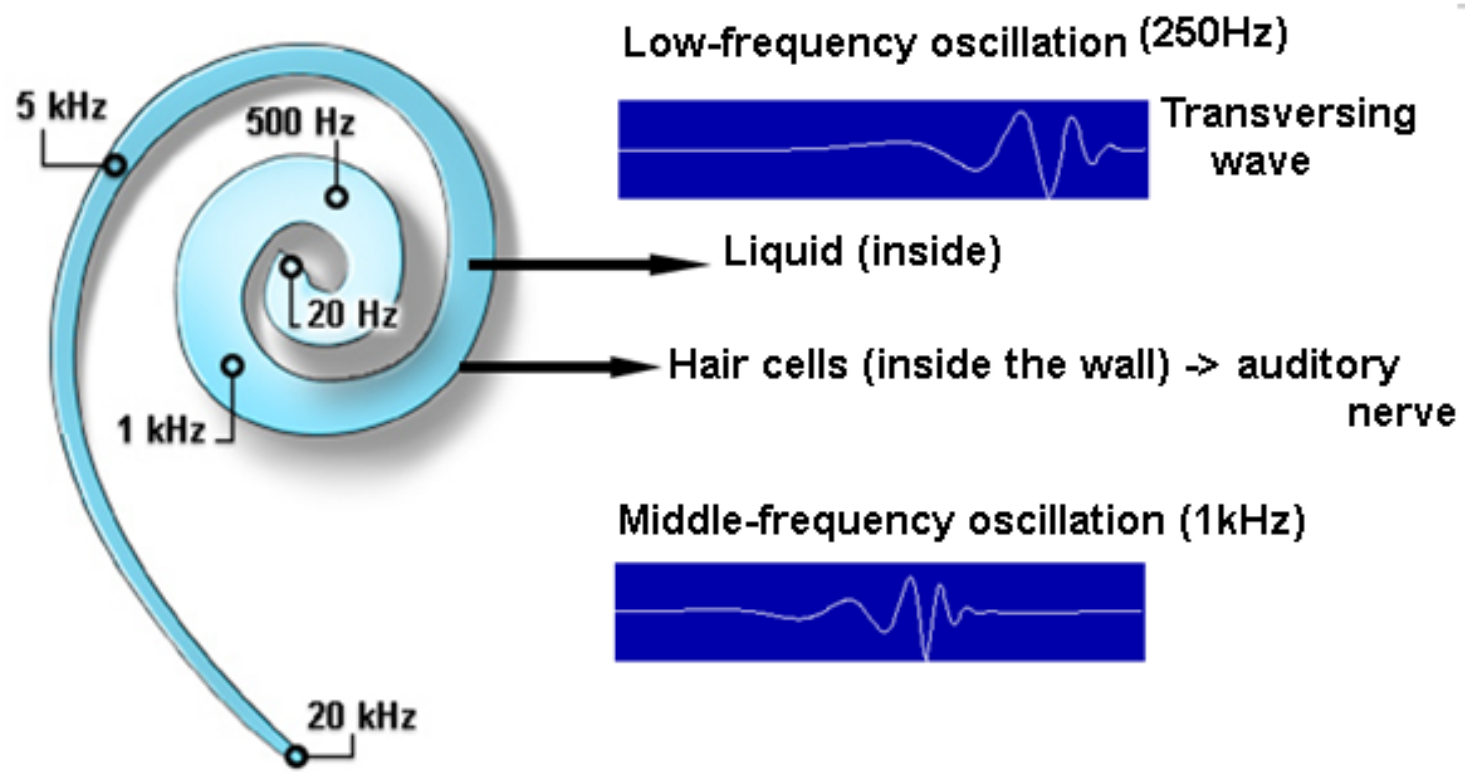
•**Outer Ear:** collects sound waves and directs them into the ear canal, which then vibrates as the sound waves hit the eardrum.

•**Middle Ear:** responsible for transmitting sound waves from the outer ear to the inner ear. The three smallest bones in the body, the malleus, incus, and stapes, work together to amplify the sound waves and transmit them to the inner ear via the oval window.

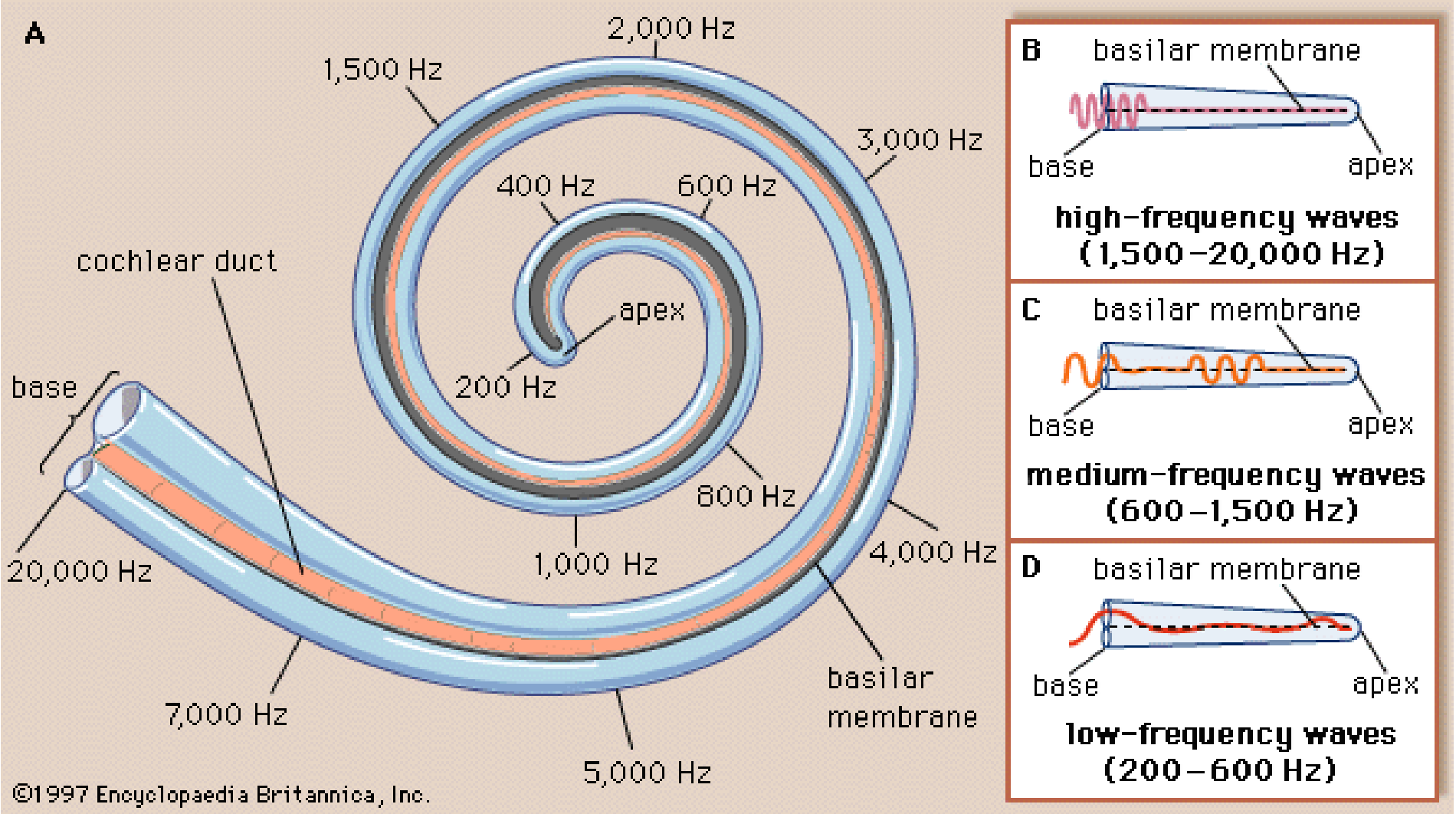
•**Inner Ear:** Converts sound waves to electrical signals for the brain to interpret and helps with balance and spatial orientation. The cochlea contains tiny hair cells that are stimulated by the vibrations of fluid inside and send electrical signals to the brain via the auditory nerve, responsible for interpreting sound.

The operation of the cochlea

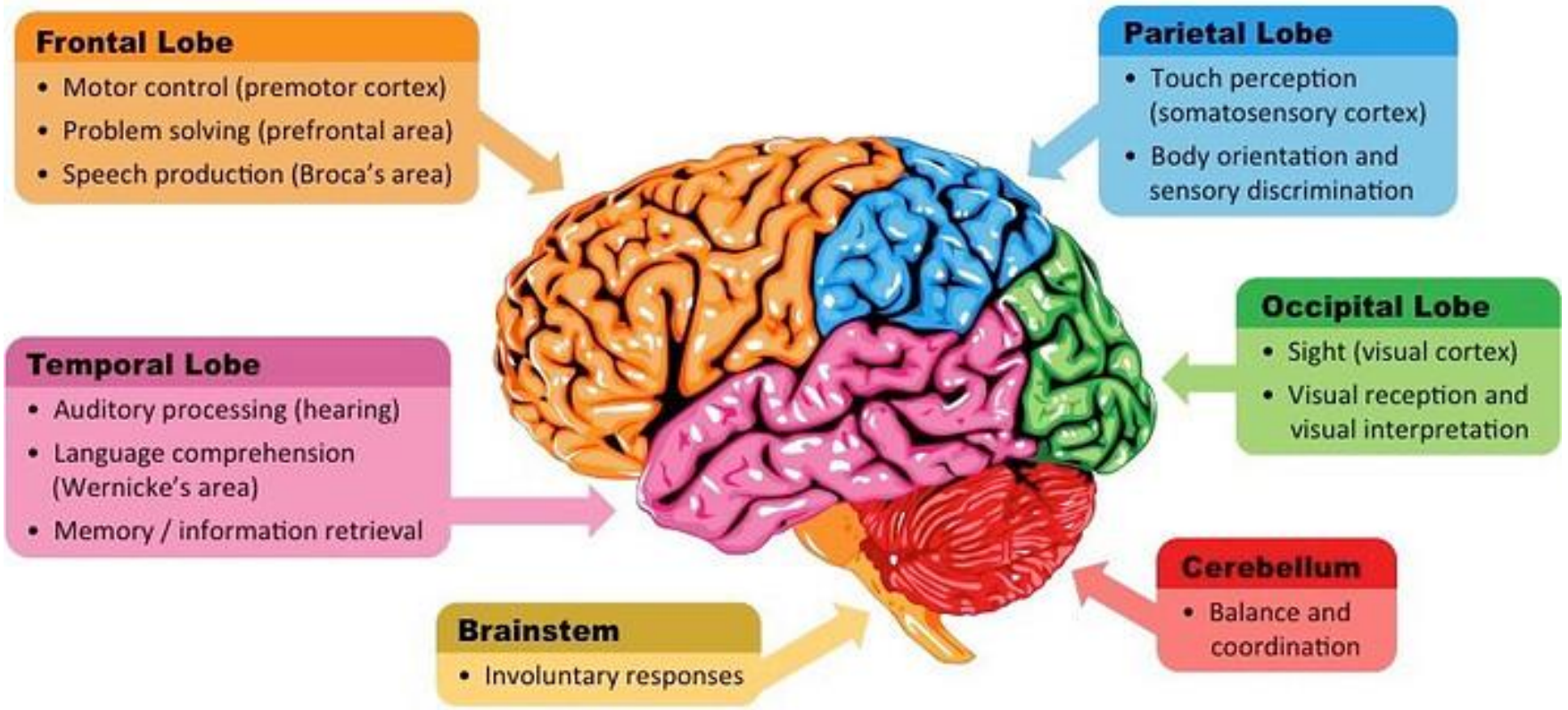
Frequency - position transformation in the cochlea (György Békésy, 1926)
Spectrogram-like transformation (nonlinear filterbank model)



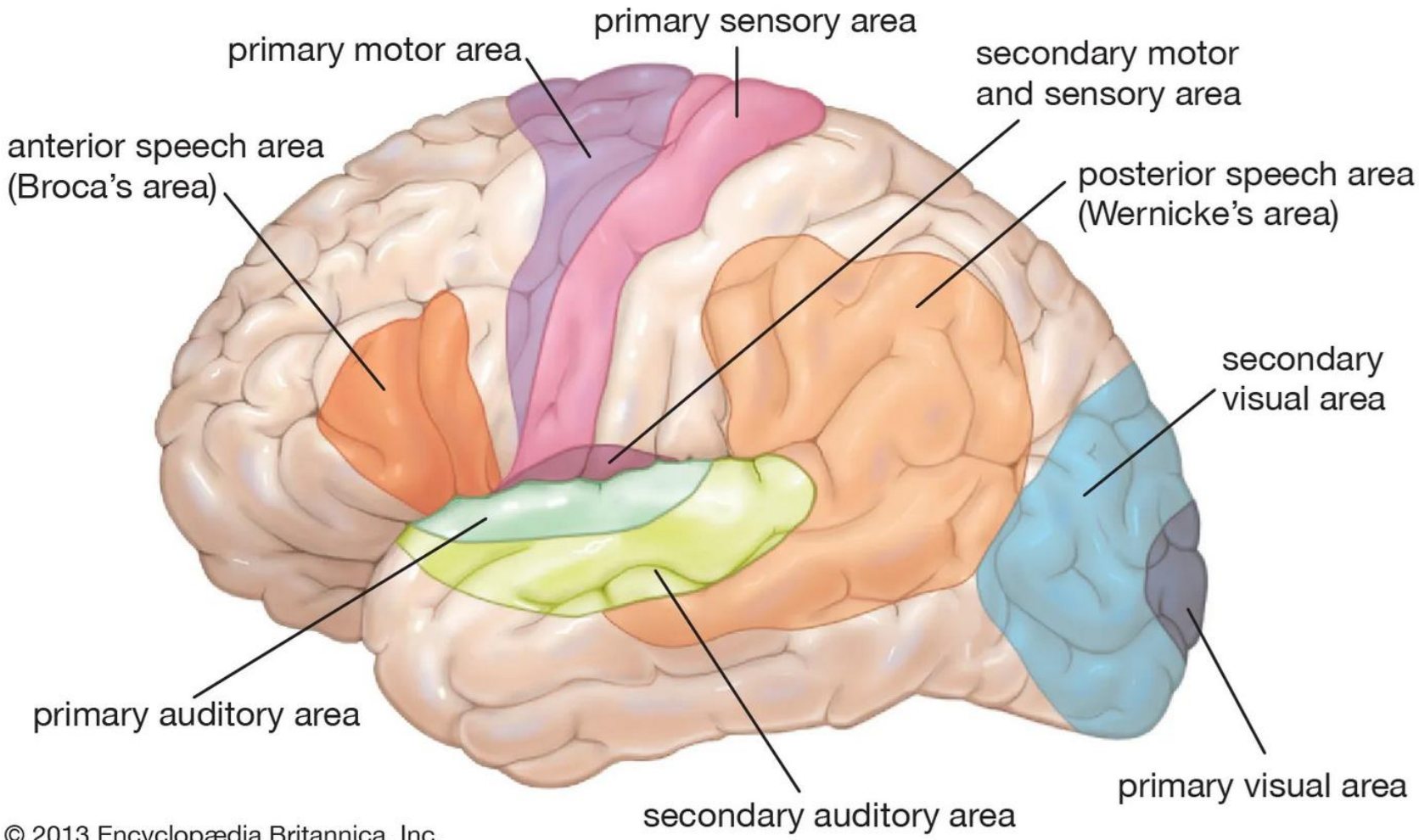
The operation of the cochlea



The anatomical structure and main functions of the brain



The anatomical structure and main functions of the brain



Source: http://opuseteducatio.hu/index.php/opusHU/article/view/173/232

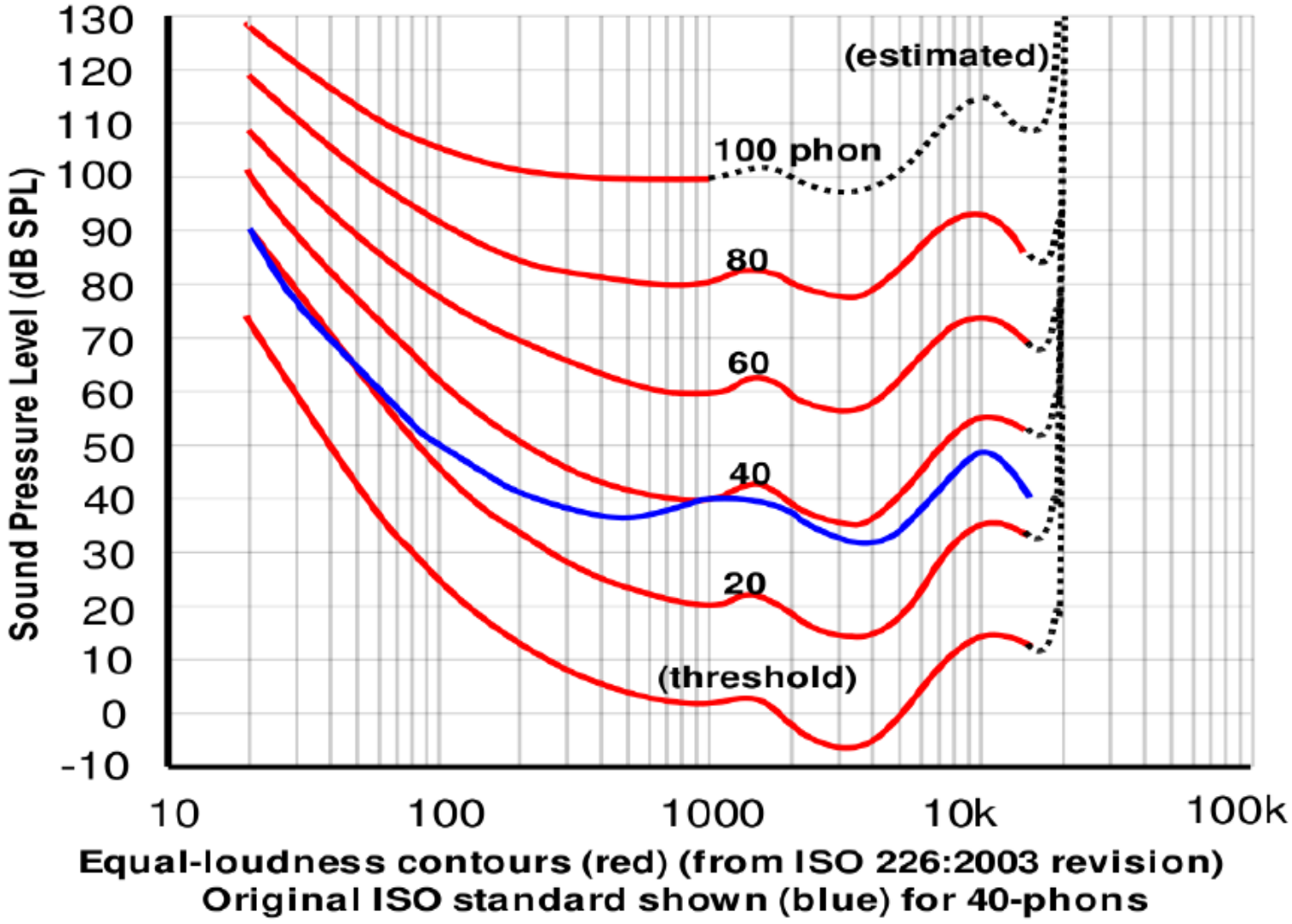
Physical and psychophysical characteristics of hearing

Physical	Psychophysical
Intensity	Loudness
Fundamental frequency (pitch, F_0) (male approx. 100Hz, female approx. 200Hz)	Melody, Intonation
Spectrum	Voice character

Equal loudness levels of a sinusoidal sound

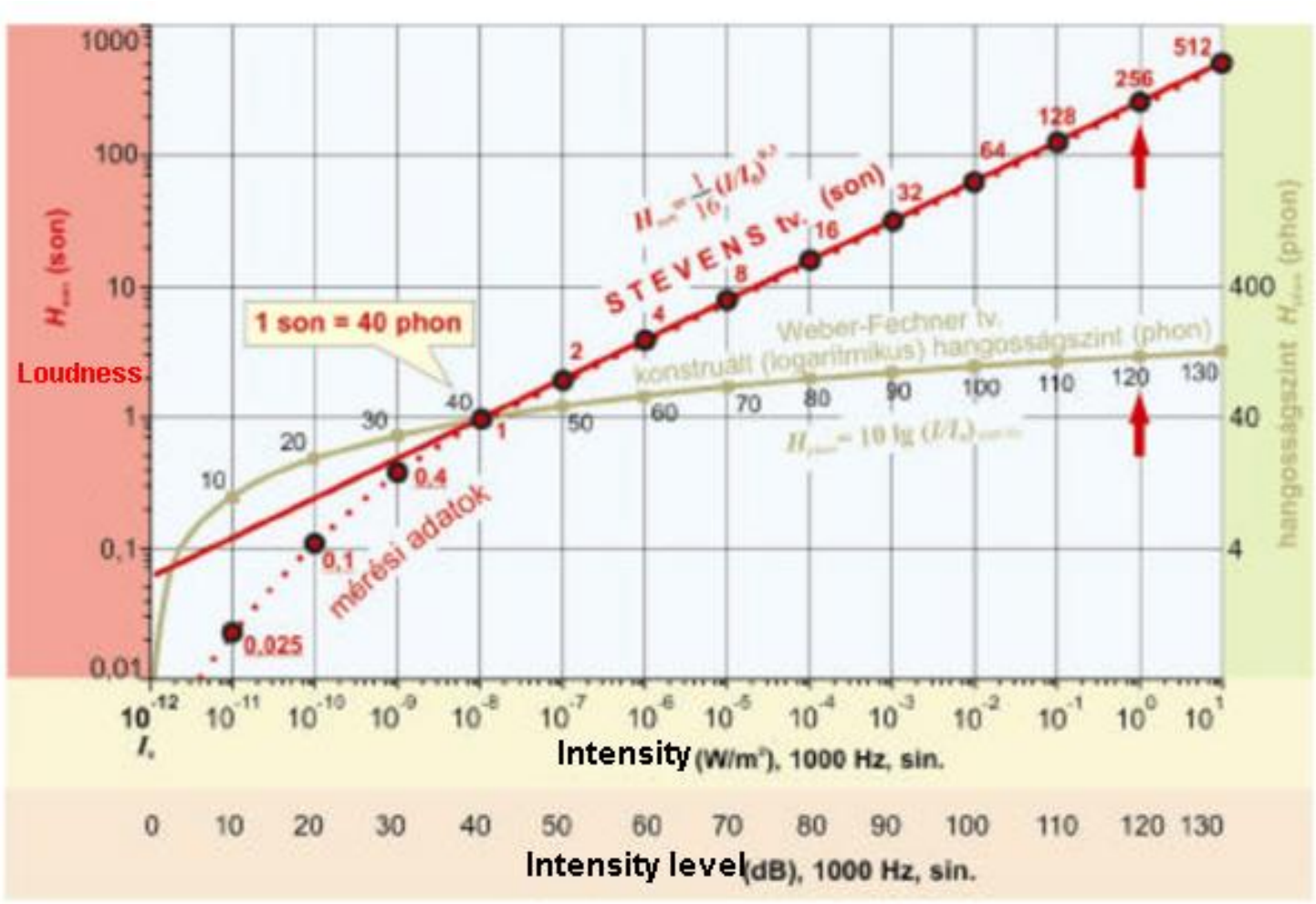
1933 Fletcher- Munson

The absolute loudness level of a sinusoidal sound measured in phon is equal to the intensity level of a 1KHz sinusoidal sound perceived as having the same loudness, measured in acoustic dB.



Source:

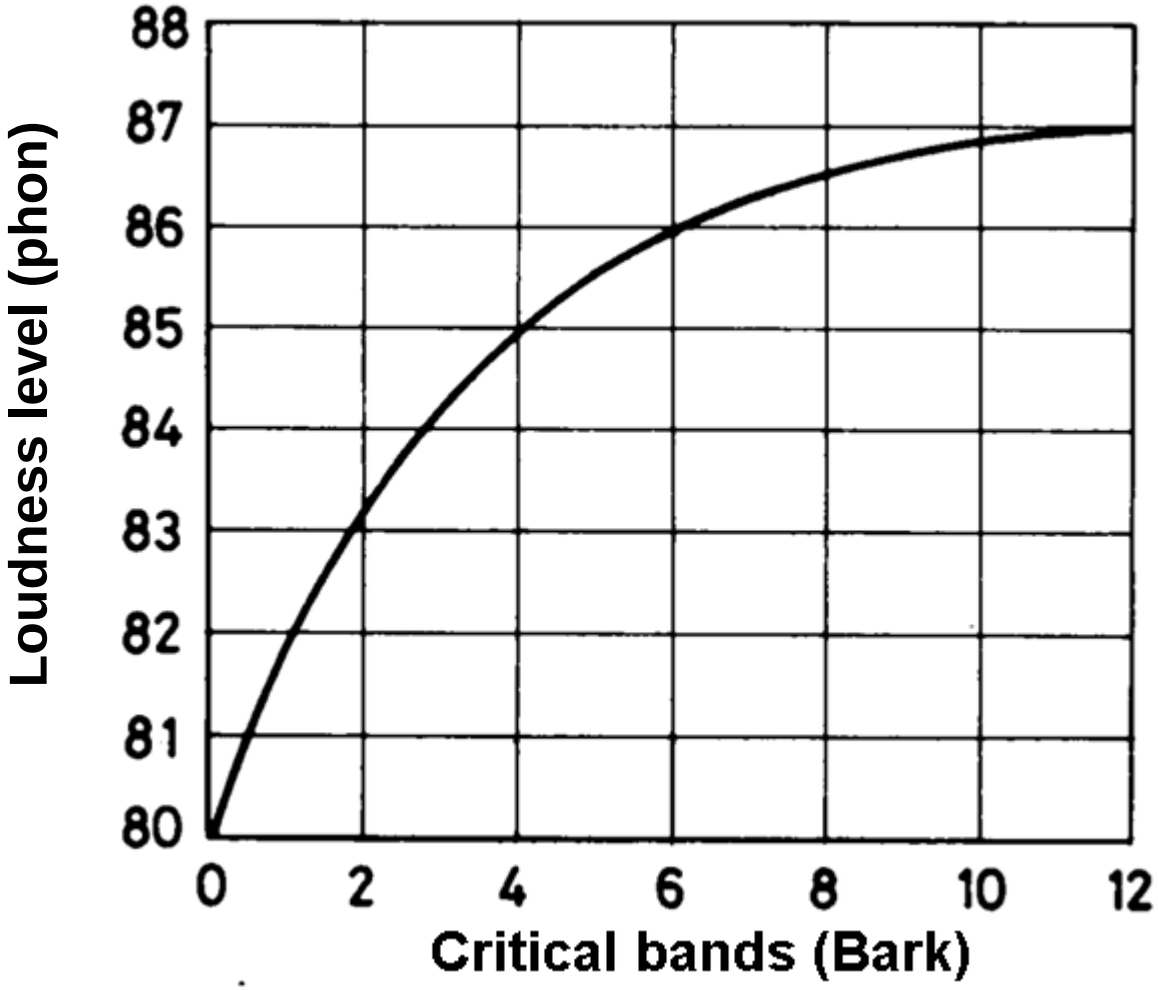
Relative loudness of sinusoidal sounds



$N = 2^{L-40[phon]/10}[son]$

$N \sim (L/40)^{0,3}$ [son]

Critical bands

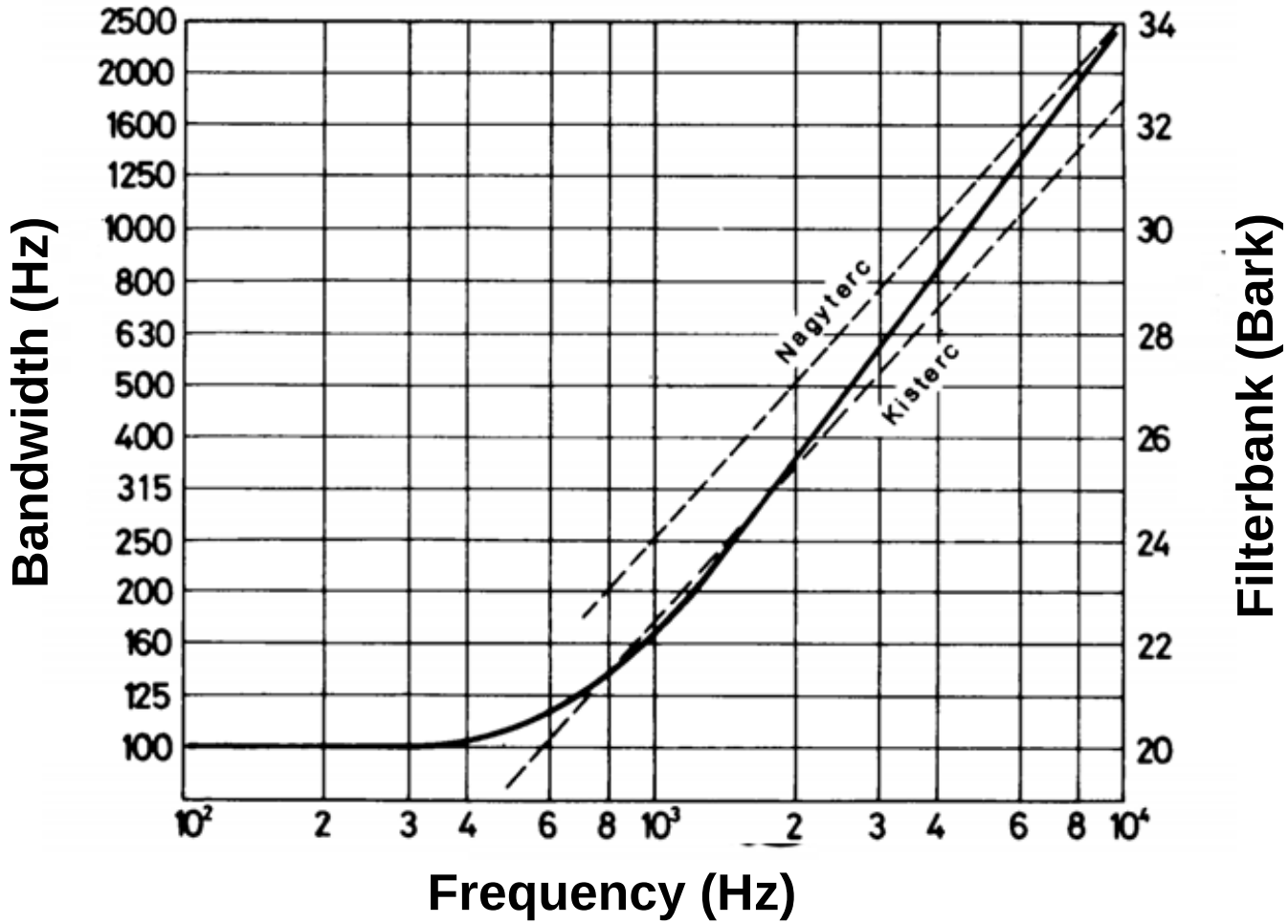


Resultant loudness level of two 77 phon sound sources as a function of frequency distance

Energy summation within the critical band, higher loudness level outside it

Source: https://last.hit.bme.hu/sites/default/files/documents/J.Granat_Hangjelek_Hallas.pdf

The critical band as a function of the mid-band frequency



Source: https://last.hit.bme.hu/sites/default/files/documents/J.Granat_Hangjelek_Hallas.pdf

Critical band ranges

Band no. [Bark]	Center frequency [Hz]	Low-High end [Hz}	Bandwidth [Hz]
1	50	20-100	80
2	150	100-200	100
3	250	200-300	100
4	350	300-400	100
5	450	400-510	110
6	570	510-630	120
7	700	630-770	140
8	840	770-920	150
9	1000	920-1080	160
10	1170	1080-1270	190
11	1370	1270-1480	210
12	1600	1480-1720	240
13	1850	1720-2000	280
14	2150	2000-2320	320
15	2500	2320-2700	380
16	2900	2700-3150	450
17	3400	3150-3700	550
18	4000	3700-4400	700
19	4800	4400-5300	900
20	5800	5300-6400	1100
21	7000	6400-7700	1300
22	8500	7700-9500	1800
23	10500	9500-12000	2500
24	13500	12000-15500	3500

Within a critical band, two narrowband white noises of equal power but different bandwidths are perceived as having the same loudness.

Source: https://last.hit.bme.hu/sites/default/files/documents/J.Granat_Hangjelek_Hallas.pdf

Frequency domain level masking