

# Einführung in die Phonetik und Phonologie

SoSe 2023

## Akustische Phonetik

Bernd Möbius / Bistra Andreeva

Sprachwissenschaft und Sprachtechnologie  
Universität des Saarlandes

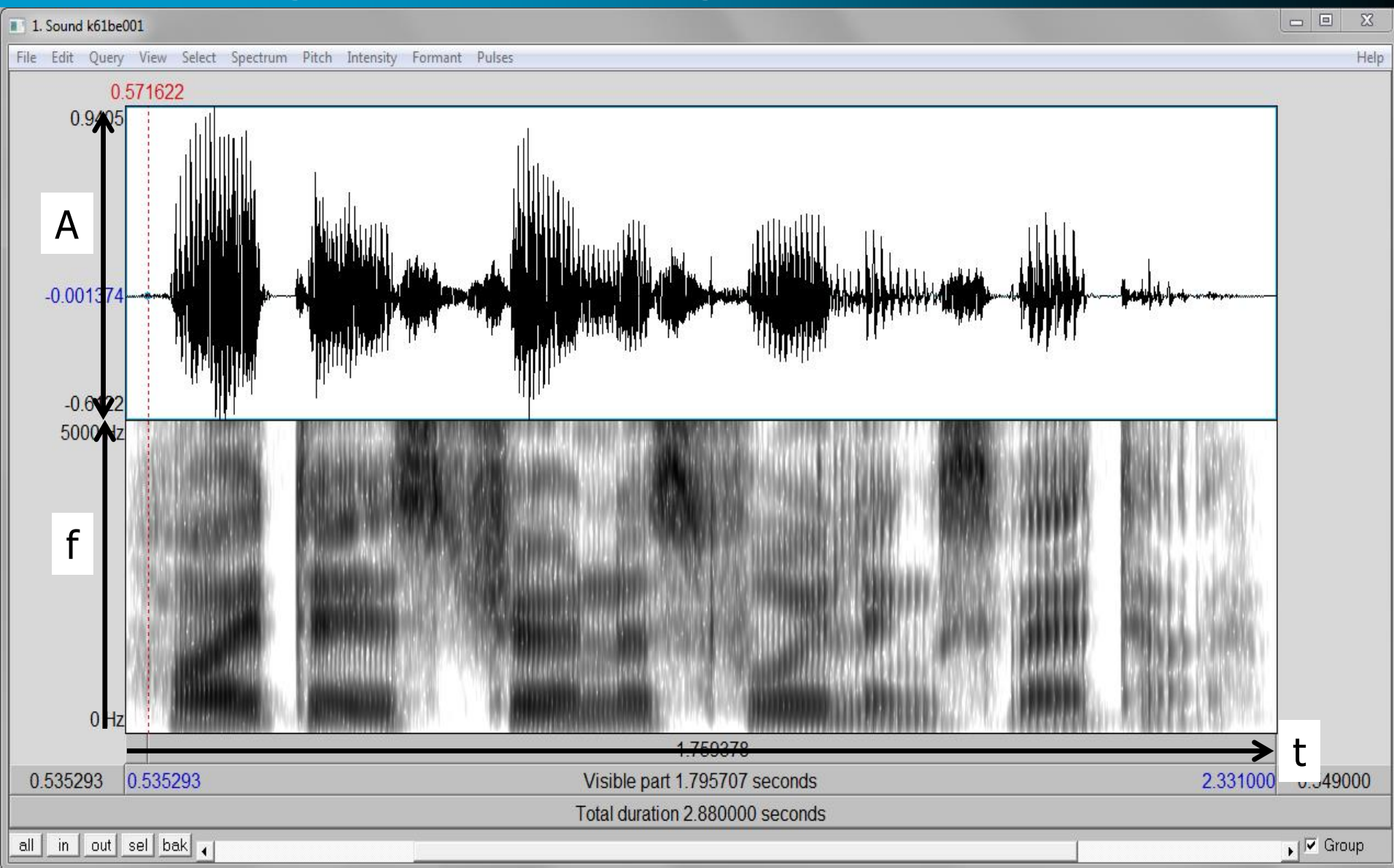


# Lautsprachliche Kommunikation

- Voraussetzungen
  - Schallerzeugung
  - Schallwahrnehmung
  - Schalleitendes Medium
- Akustische Basiseigenschaften von Sprachlauten
  - Frequenzen innerhalb des menschlichen Wahrnehmungsbereichs (20 – 20000 Hz)
  - Amplitude: Auslenkung einer Schwingung → wahrgenommene Lautstärke
  - Dauer: wahrnehmbare Mindestdauer
  - Klangfarbe (Timbre)

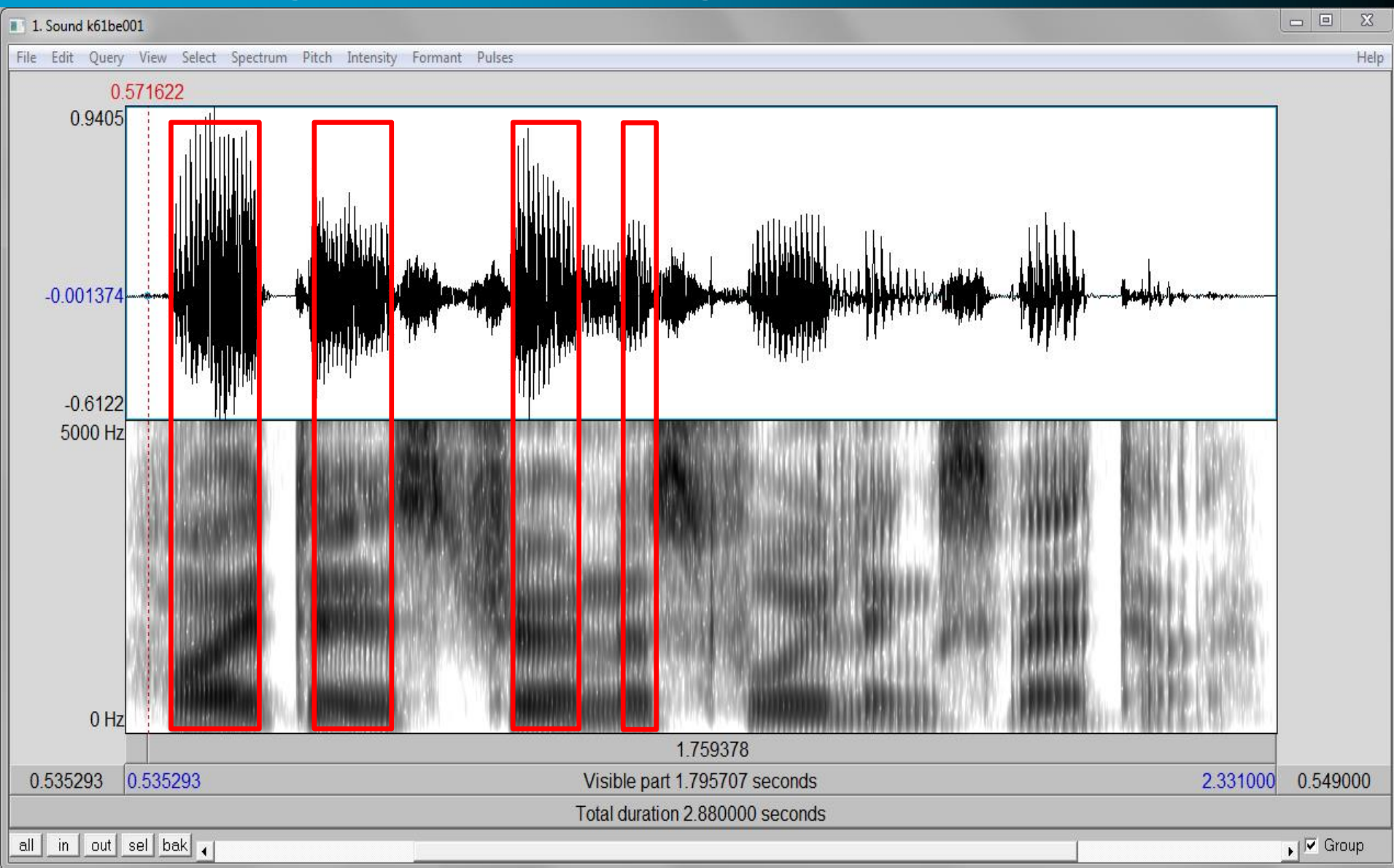


# Sprachsignal und Spektrogramm



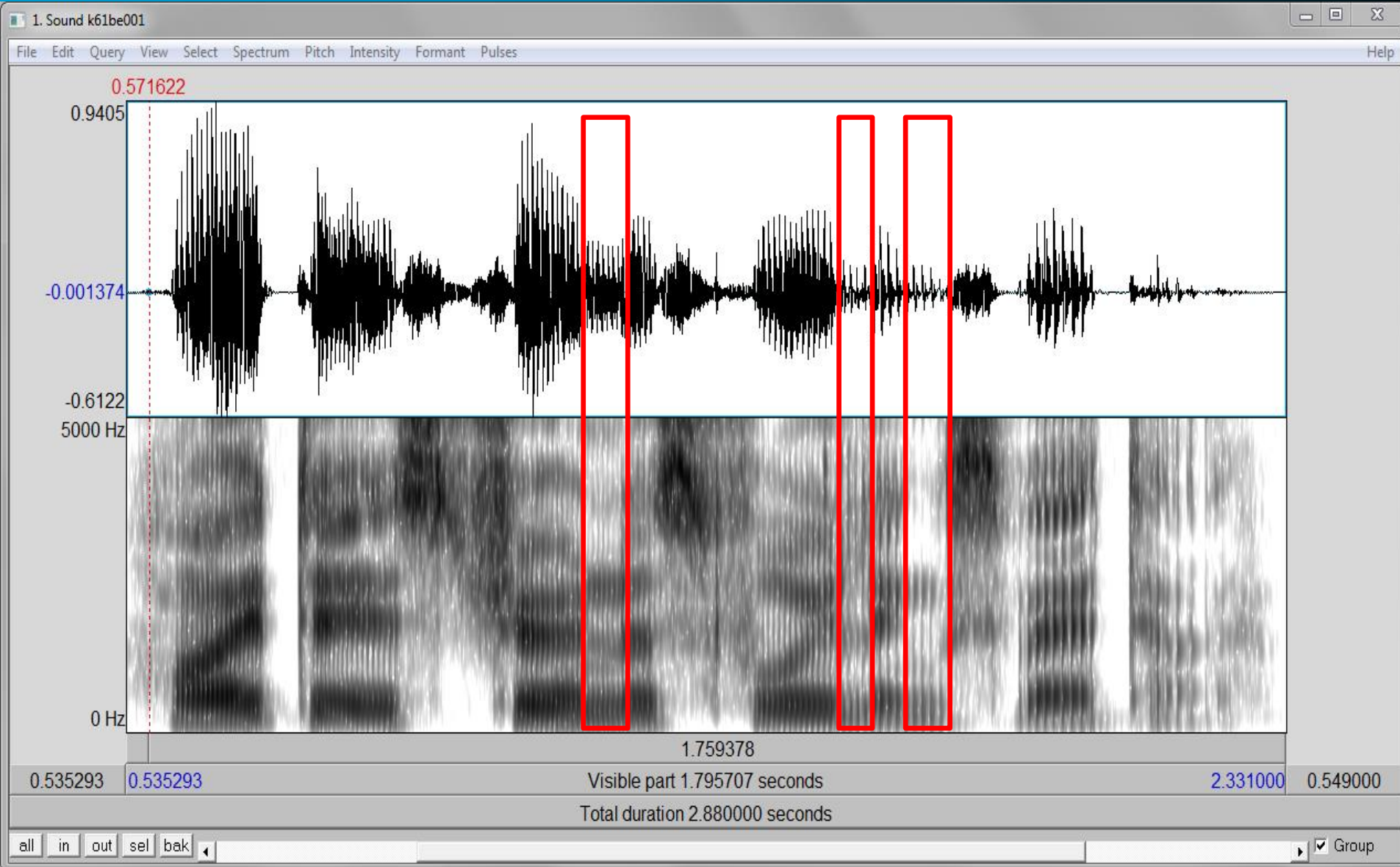
"Heute ist schönes Frühlingswetter."

# Sprachsignal und Spektrogramm: Vokale



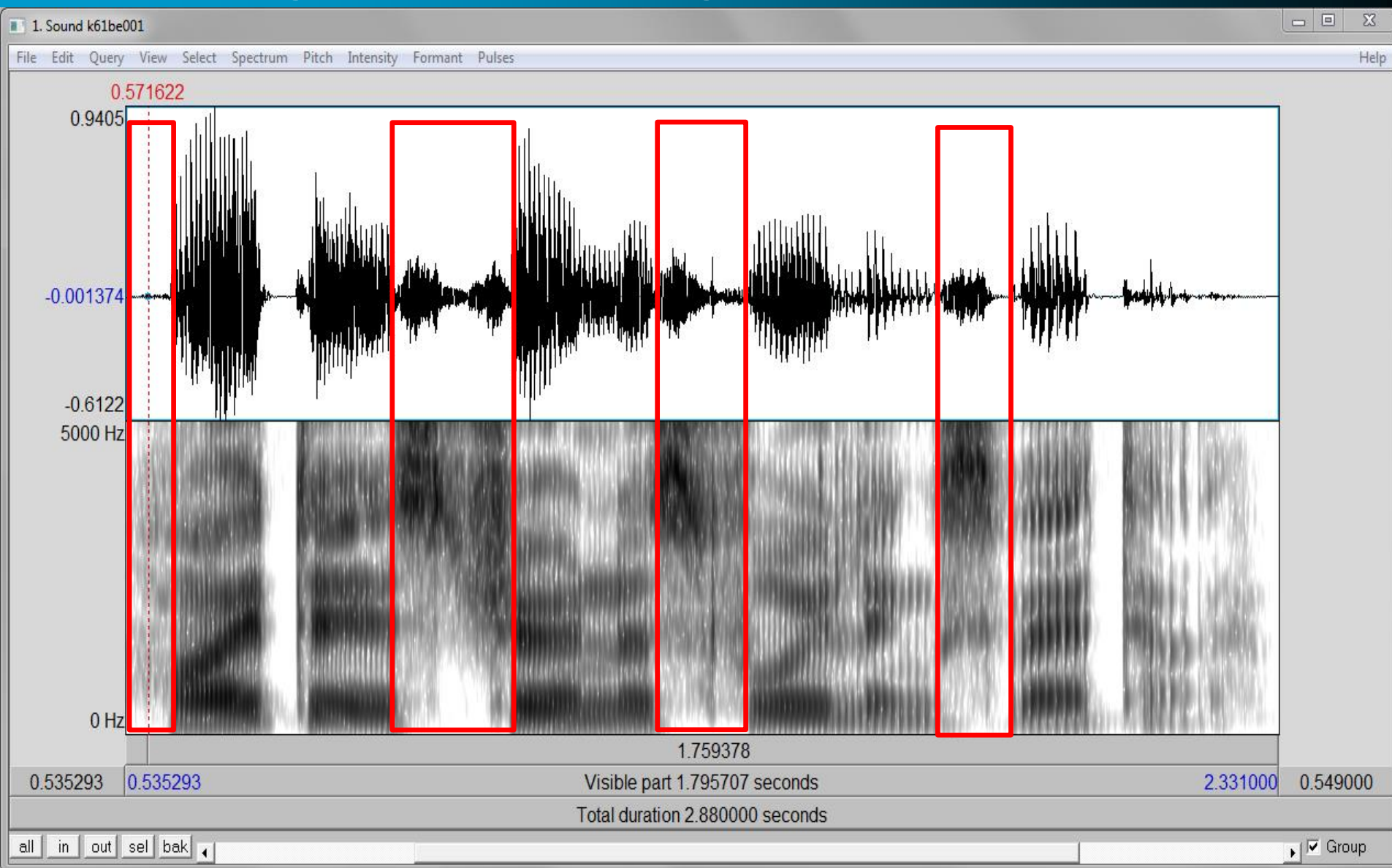
"Heute ist schönes Frühlingswetter."

# Sprachsignal und Spektrogramm: Nasale



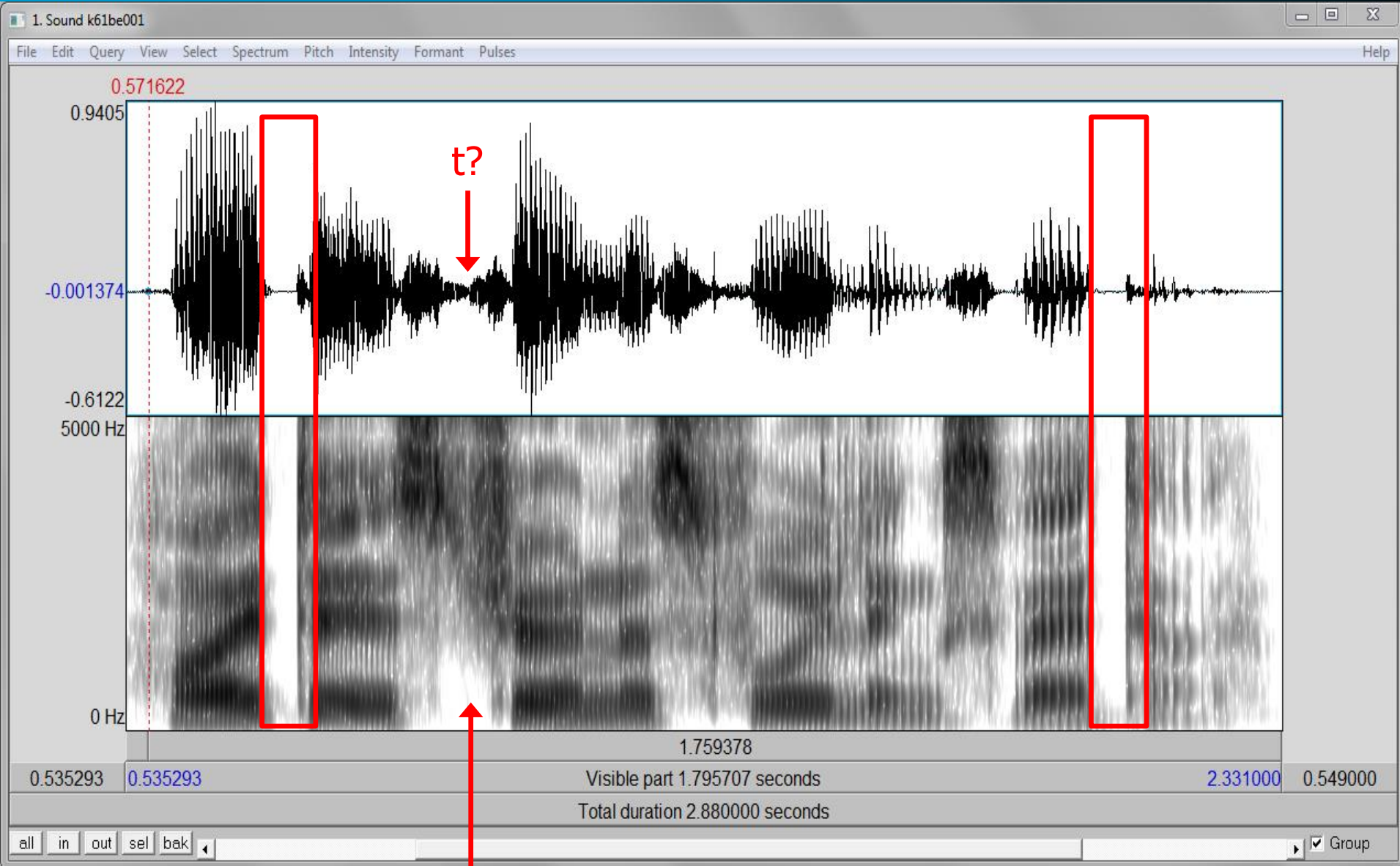
"Heute ist schönes Frühlingswetter."

# Sprachsignal und Spektrogramm: Frikative



"Heute is schönes Frühlingswetter."

# Sprachsignal und Spektrogramm: Plosive



"Heute is(t) schönes Frühlingswetter."

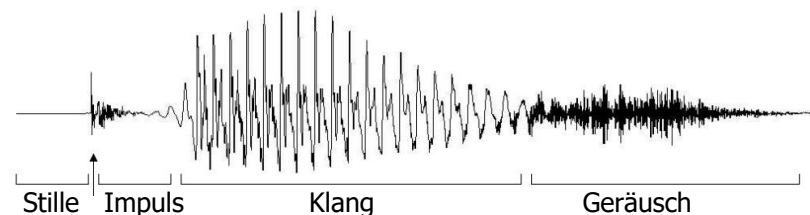
# Sprachlaute...: stimmhafte Frikative



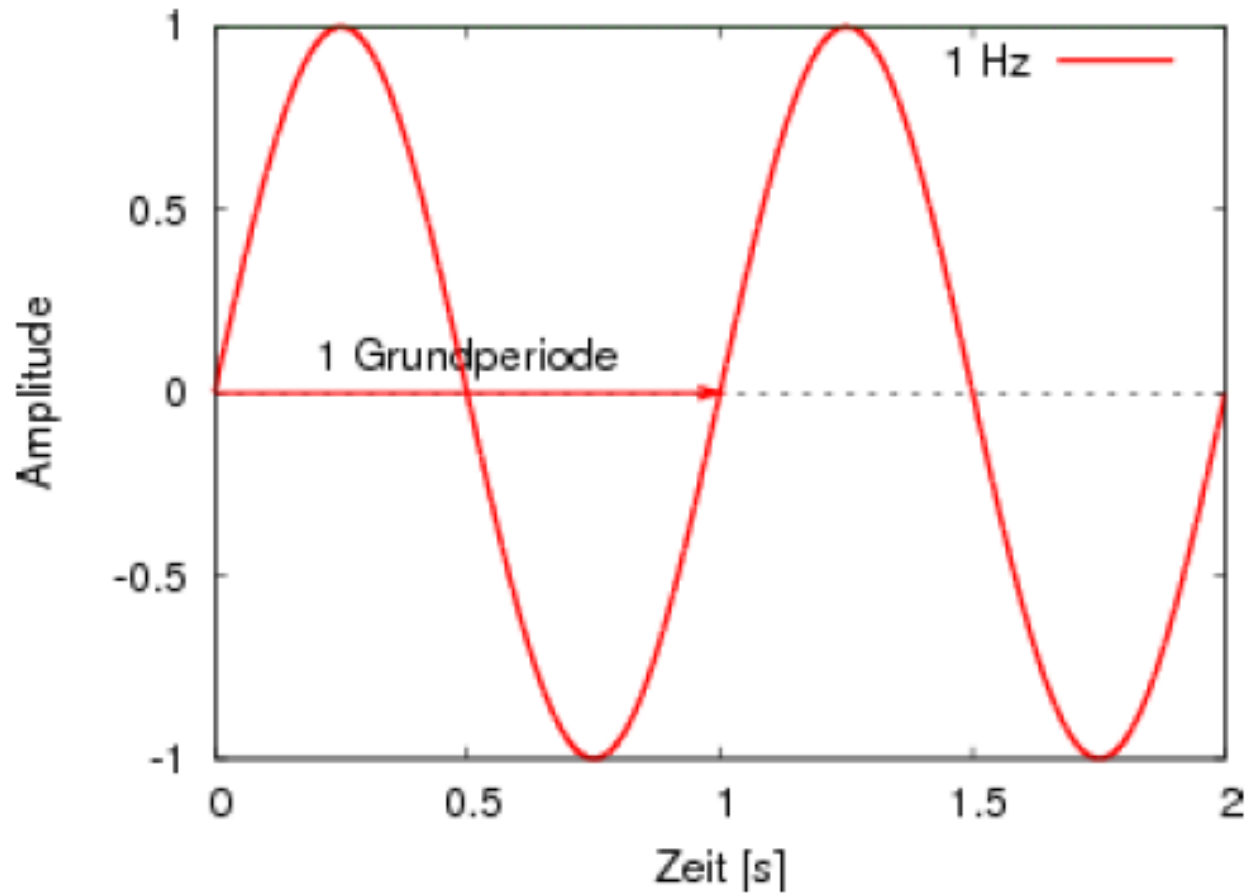
"Heute ist schönes Frühlingswetter."

# Signalformen und Lautklassen

- Grundlegende Formen des Sprachsignals
  - quasi-periodische Signalformen: Klang (*sonority*)
    - Vokale
    - Sonoranten (Approximanten, Glides, Nasale, Liquide)
  - stochastische Signalformen: Rauschen (*noise*)
    - Frikative
    - Aspiration von Plosiven
  - gemischte Anregung (*mixed excitation*): stimmhaftes Rauschen
    - stimmhafte Frikative
  - transiente Signalformen: Impuls/Knall
    - Plosivlösungen



# Einfache Schwingungen



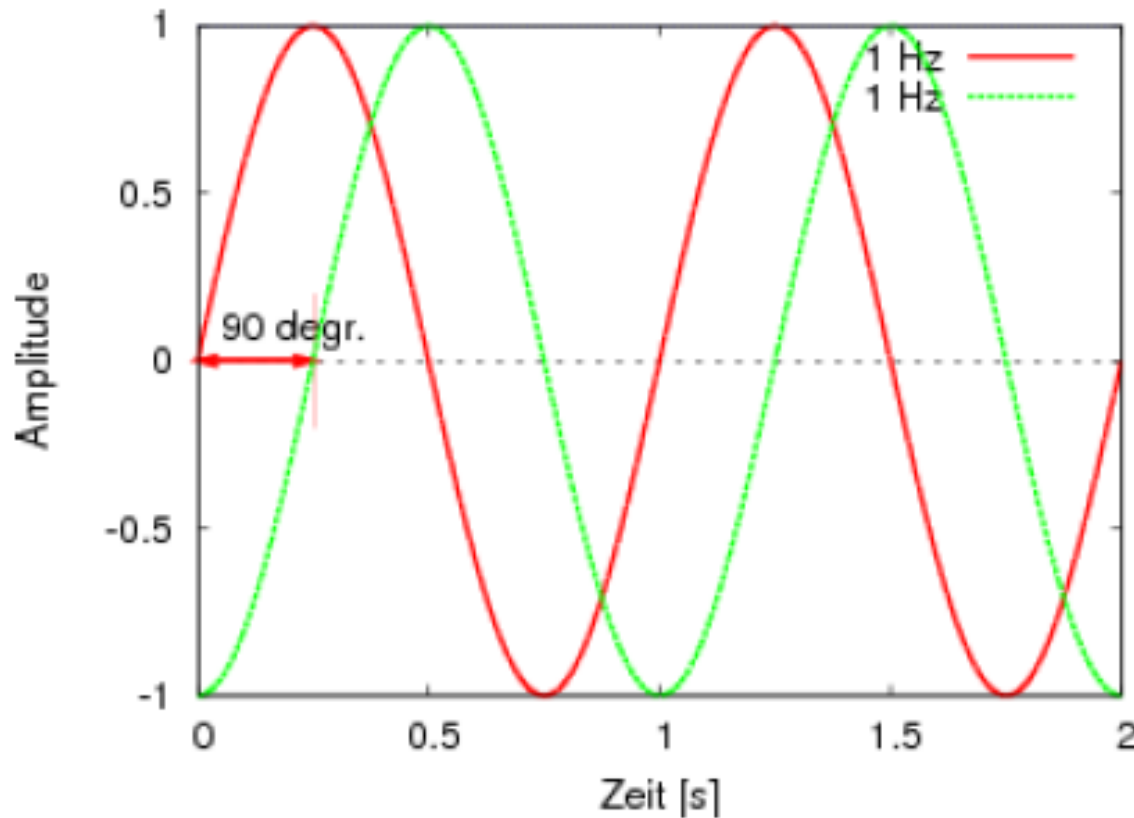
# Einfache Schwingungen

- Einfache periodische Schwingung (Sinusschwingungen)
  - zyklisch wiederkehrende, einfache Schwingungsmuster
  - bestimmt durch
    - Grundperiode  $T_0$
    - Amplitude  $A$
    - Phase  $\Phi$
- Grundfrequenz [Hz]:  $1 / \text{Grundperiode [s]}$

$$F_0 = 1 / T_0$$

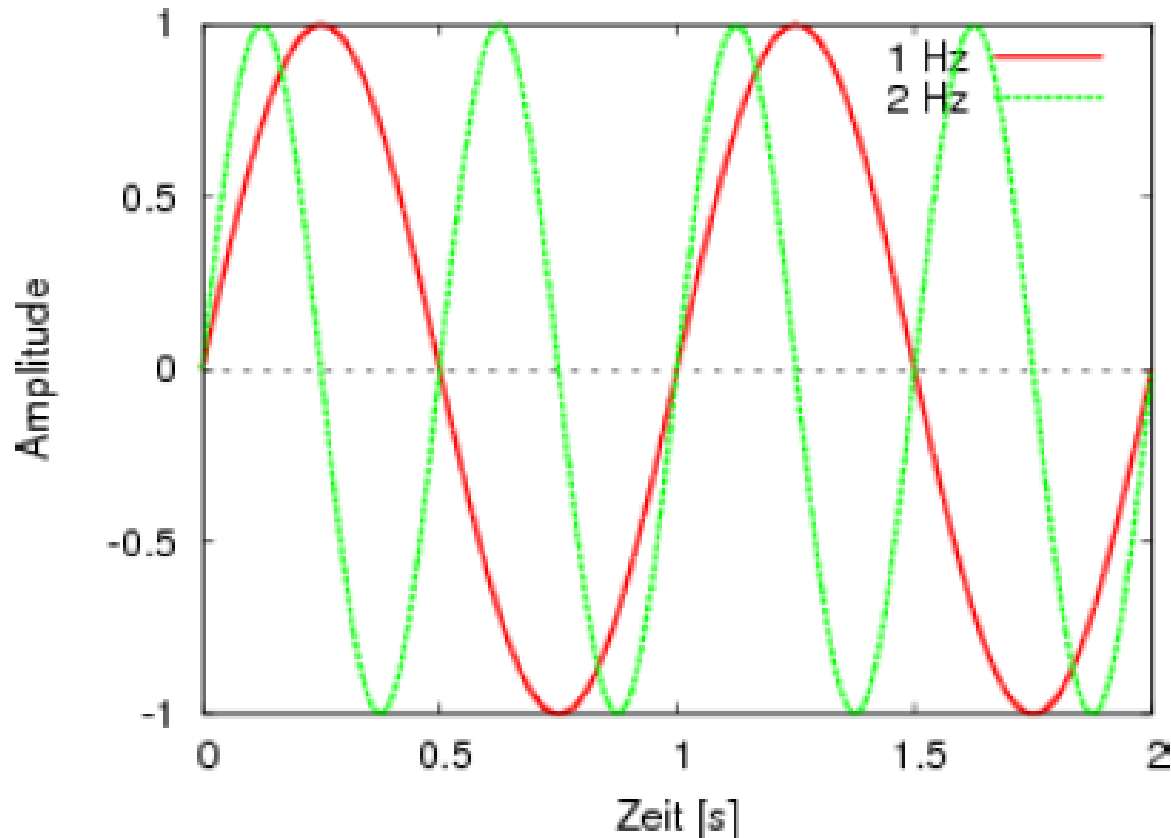
# Phasenbeziehung

- Zwei Sinusschwingungen mit gleicher Frequenz und Amplitude, aber zeitlich versetzten Minima, Maxima und Nulldurchgängen  
→ Phasenverschiebung (hier: Phasenwinkel  $90^\circ$ )



# Frequenzunterschiede

- Zwei Sinusschwingungen mit gleicher Amplitude und Phase, aber unterschiedlicher Frequenz (hier: 1 bzw. 2 Hz)

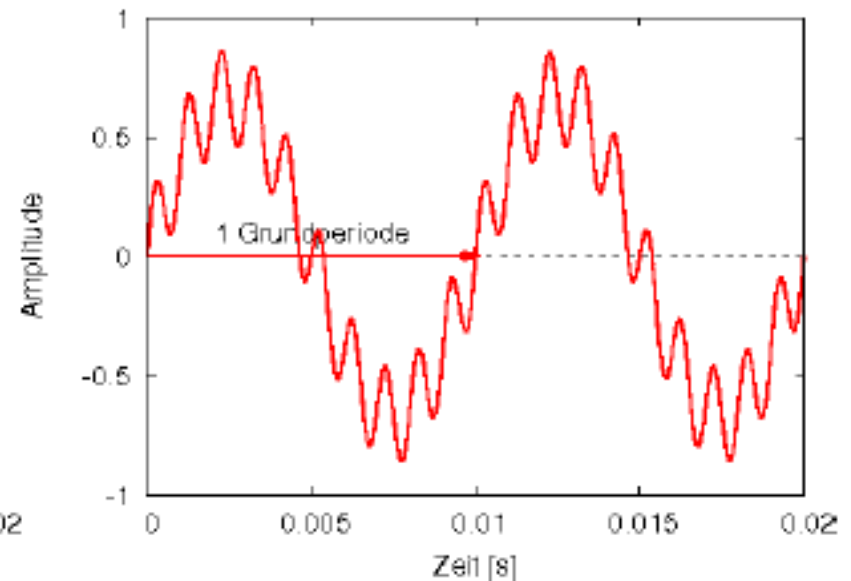
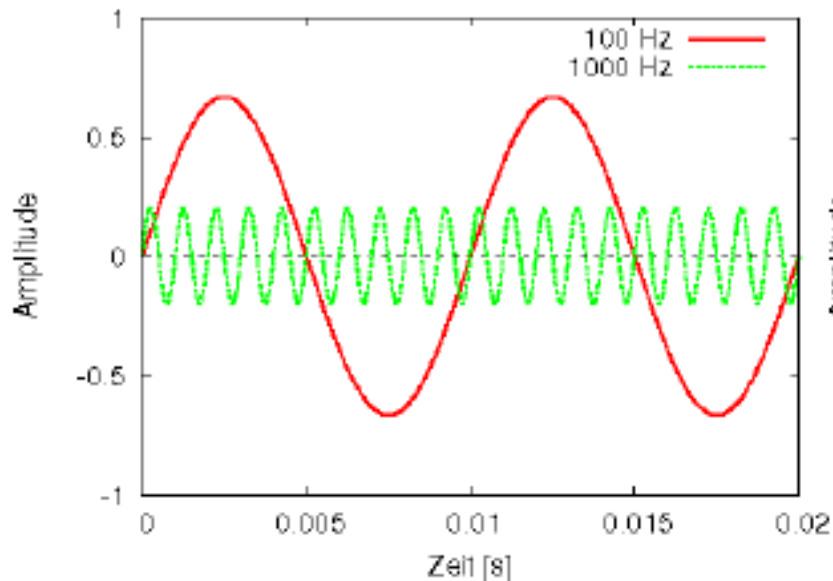


# Komplexe Schwingungen

- Komplexe periodische Schwingungen
  - zyklisch wiederkehrende Schwingungsmuster
  - aus mindestens zwei Sinusschwingungen zusammengesetzt
  - Grundfrequenz =  $1 / (\text{komplexe Grundperiode})$
- Form der resultierenden komplexen Schwingung hängt von der Frequenz-, Amplituden- und Phasenrelation der Komponenten ab

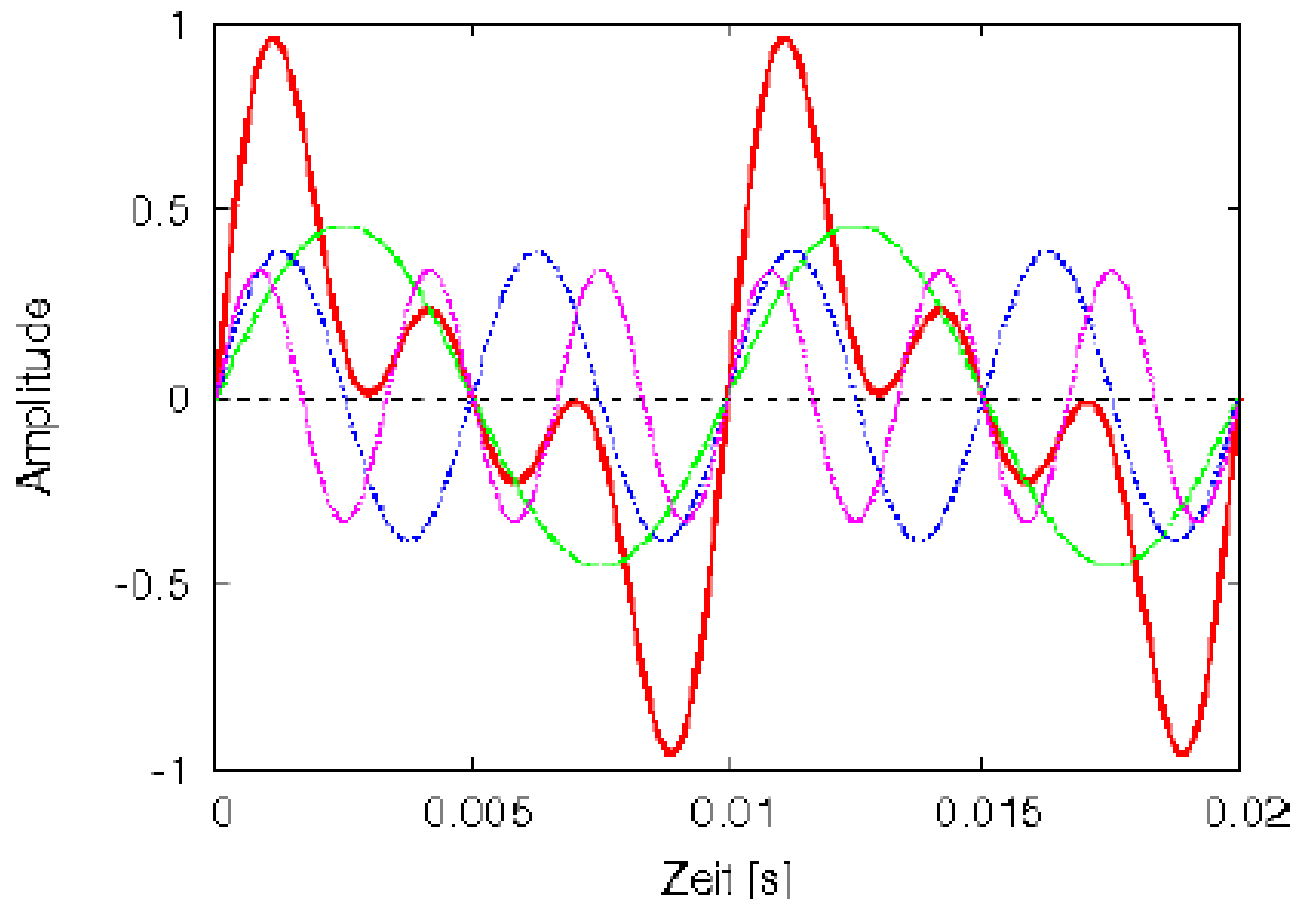
# Komplexe Schwingungen

- Komplexe Schwingung: 2 Komponenten
  - zwei Sinusschwingungen (100 Hz, 1000 Hz) mit gleicher Phase und unterschiedlicher Amplitude (links)
  - komplexe Schwingung (rechts) resultiert aus der Addition der beiden Komponenten
- $F_0 = 100$  Hz



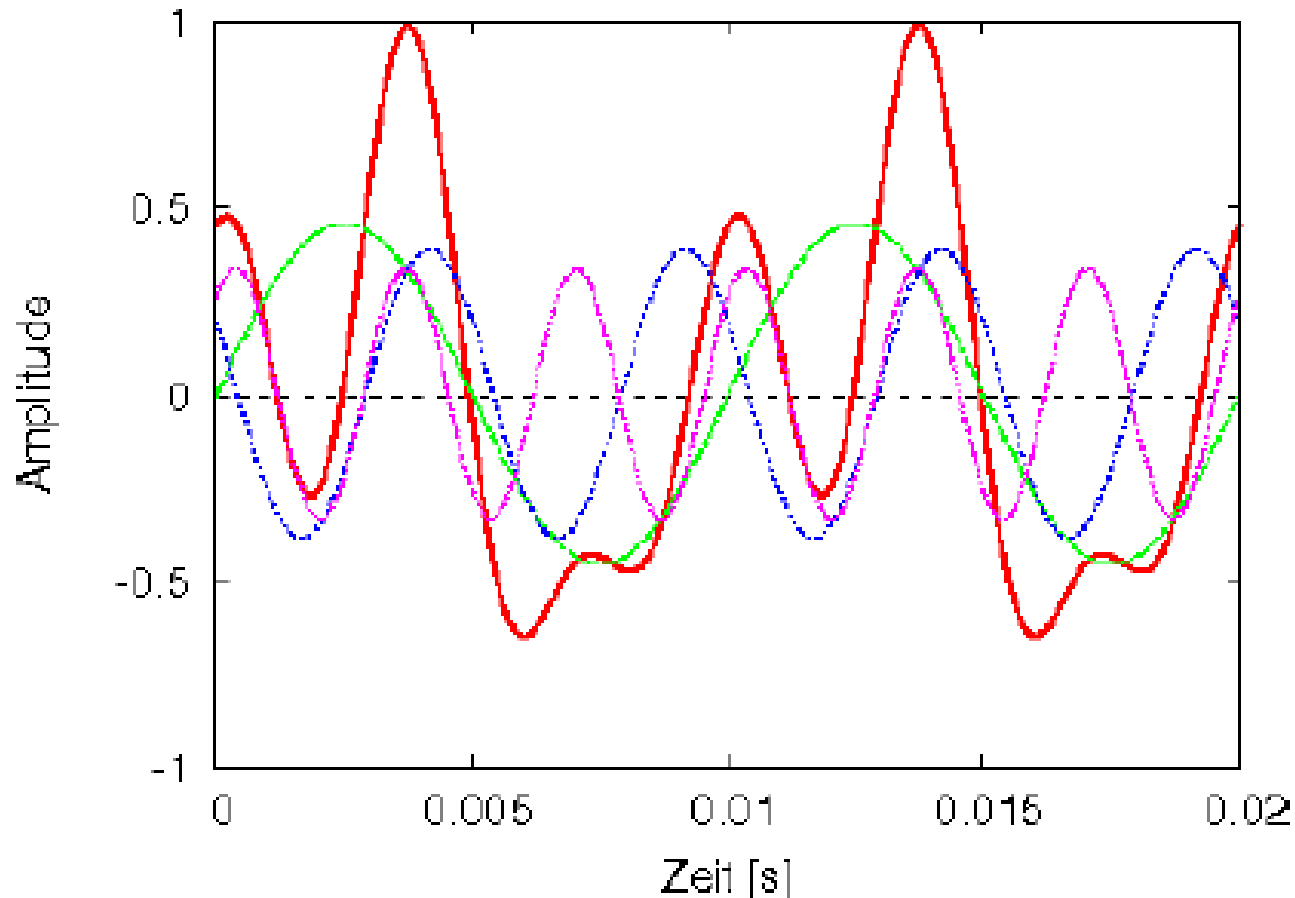
# Komplexe Schwingungen

- Komplexe Schwingung (rot): 5 Komponenten
  - 5 phasengleiche Sinusschwingungen (100, 200, 300, 400, 500 Hz)
  - dargestellt: nur 3 niederfrequenteste Komponenten



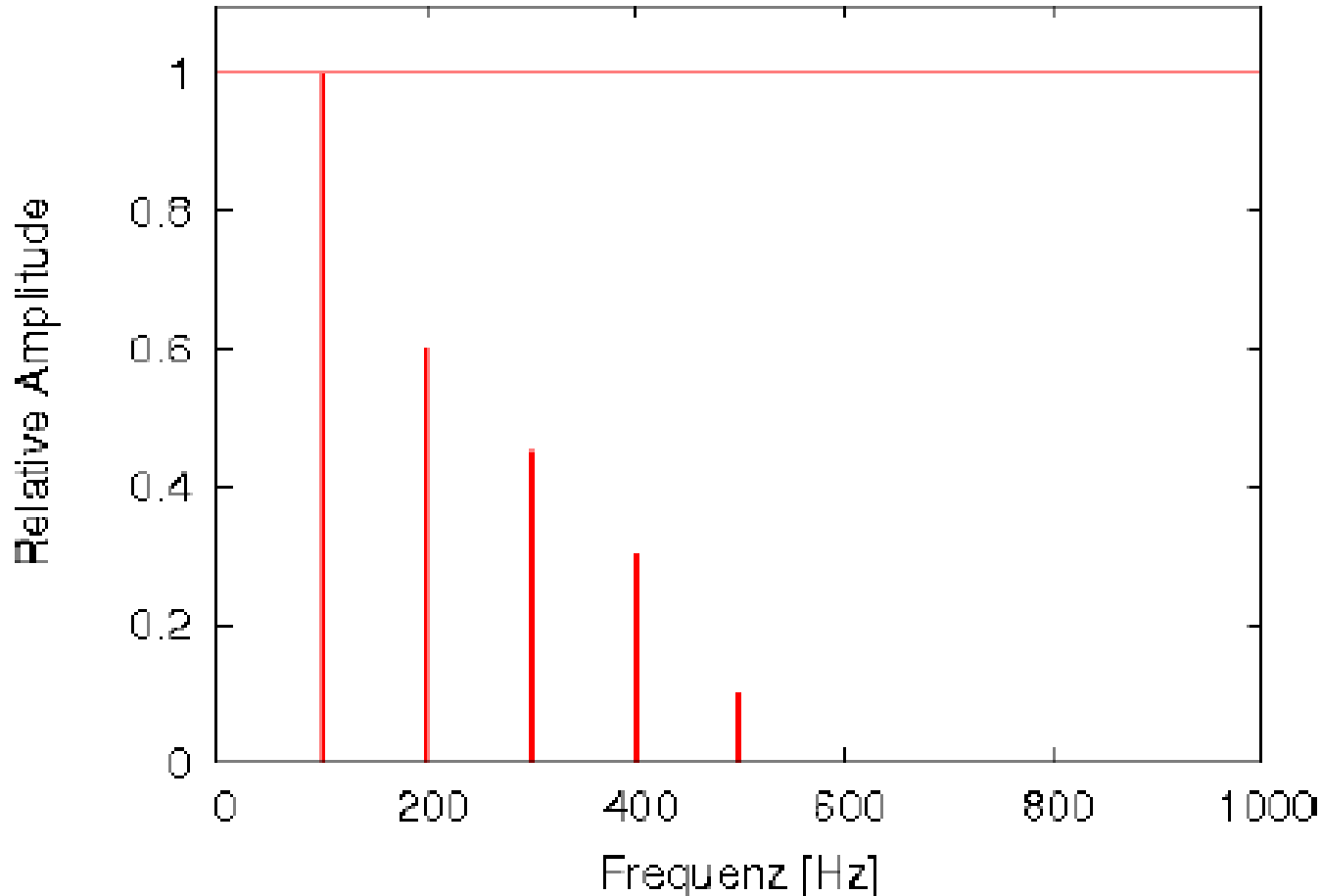
# Komplexe Schwingungen

- Komplexe Schwingung (rot): 5 Komponenten
  - 5 phasenverschobene Sinusschwingungen (100, 200, 300, 400, 500 Hz)
  - dargestellt: nur 3 niederfrequenteste Komponenten



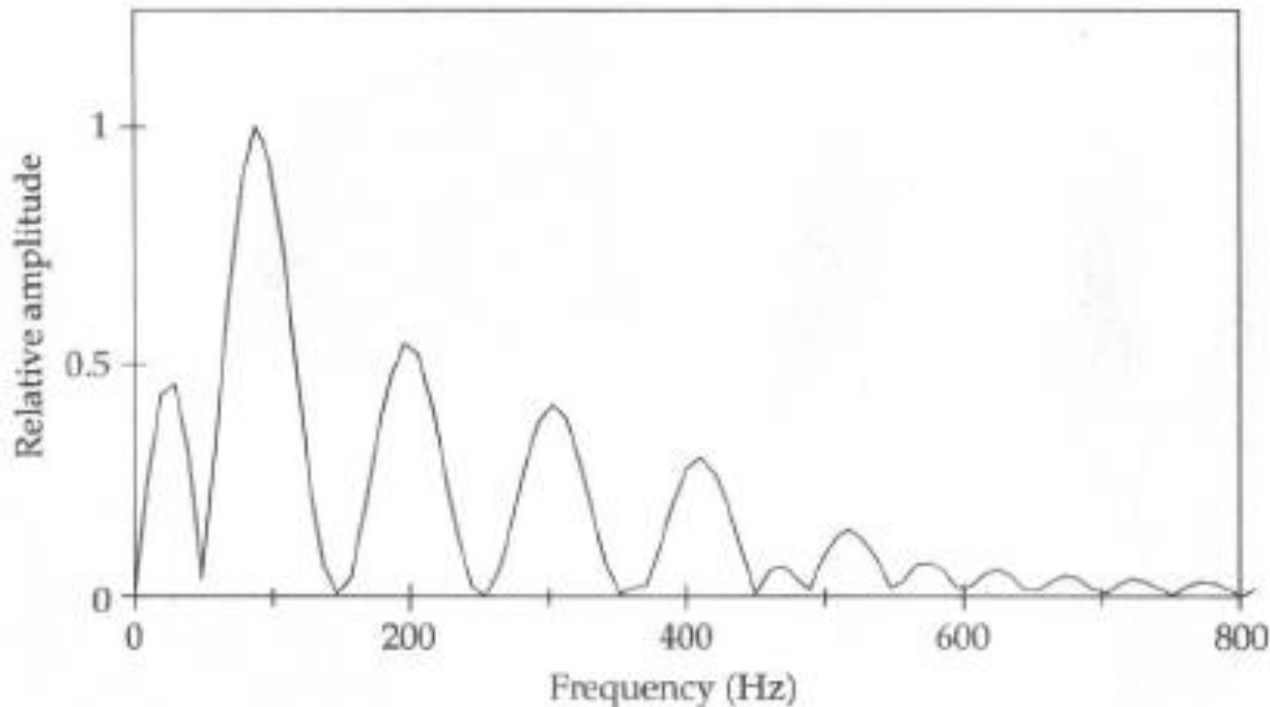
# Leistungsspektrum

- Leistungsspektrum (Amplituden über Frequenzen) der komplexen Schwingung aus 5 Komponenten



# Fourieranalyse

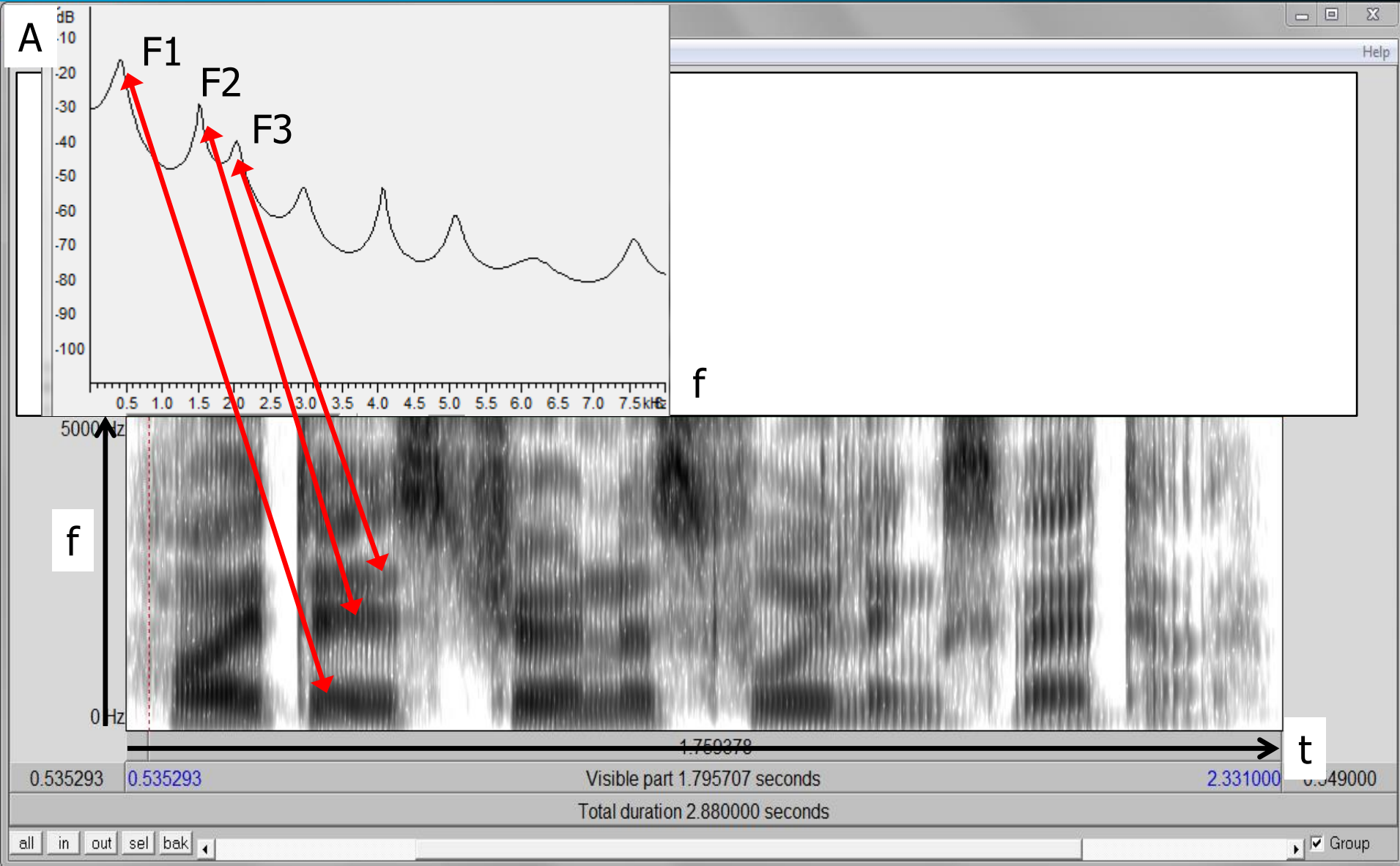
- Fouriers Theorem: Jede komplexe Schwingung kann zerlegt werden in eine Reihe von Sinusschwingungen mit jeweils eigener Frequenz, Amplitude und Phase.



- Fourieranalyse: Leistungsspektrum der komplexen Schwingung aus 5 Komponenten

- Unterschiede zwischen Ergebnis der Fourieranalyse und idealisiertem Leistungsspektrum:
  - breitere Gipfel
  - zusätzliche Gipfel
- Ursachen für diese Unterschiede:
  - Fourieranalyse nimmt unendlich langes Signal an; Analyse aber nur über 2 Grundperioden (quasi-periodisches Signal)
  - Unterschiede zwischen analogem und digitalem Signal

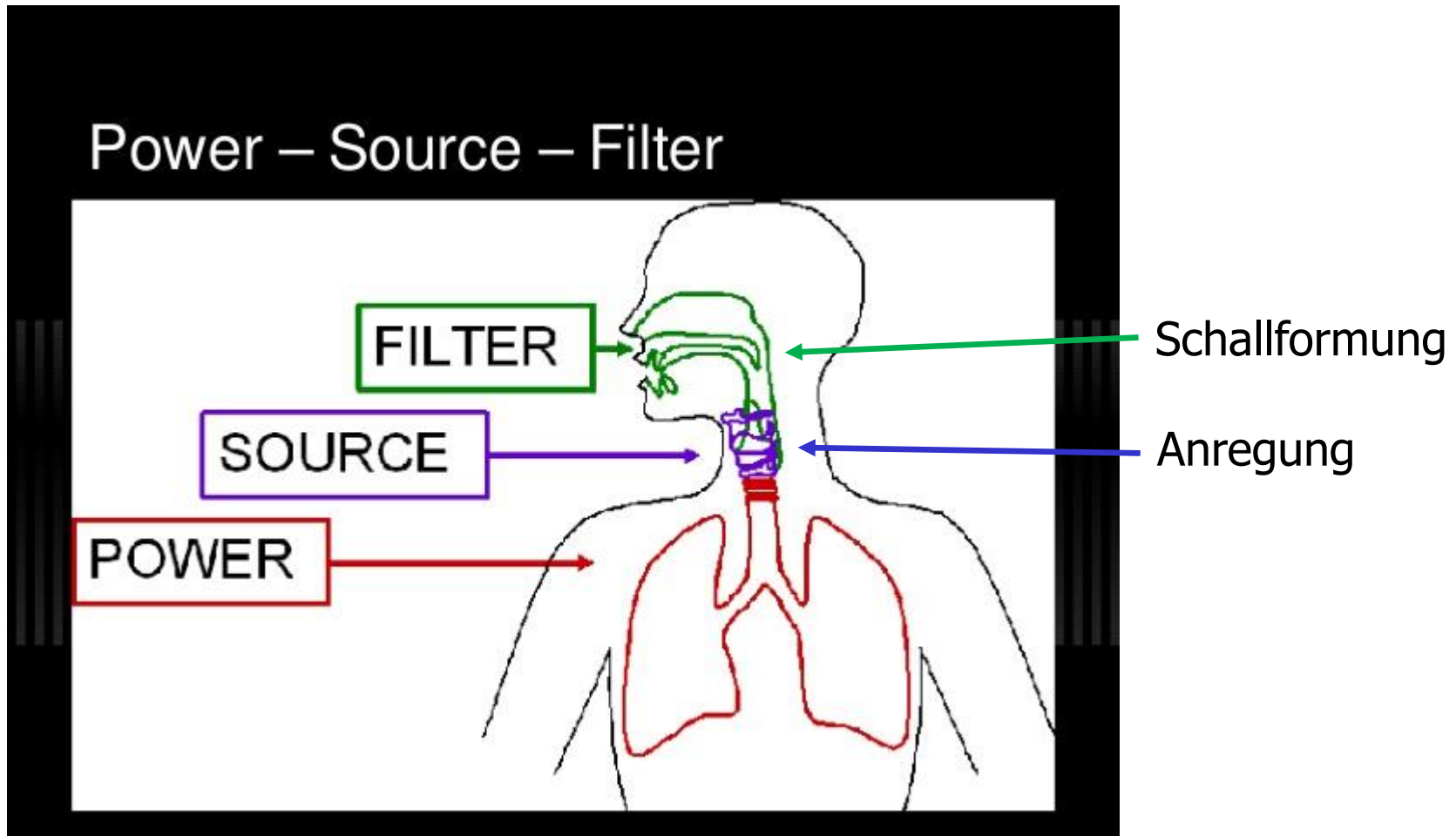
# Sprachsignal und Spektrogramm



# Formanten

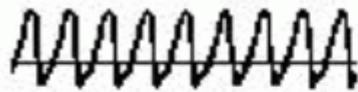
- Spektrale Gipfel (Energimaxima): **Formanten (F1, F2, ..., Fn)**
- Formanten entstehen infolge selektiver Verstärkung bestimmter Frequenzen, entsprechend der **Resonanz**charakteristika des menschlichen Vokaltrakts (und ggf. Nasaltrakts).
- Die Unterscheidung zwischen **Stimmquelle** (*Anregung*) und *Schallformung* im Vokaltrakt (akustisches **Filter**) führt zum **Quelle-Filter-Modell** der Sprachproduktion.
- Literatur:
  - Gunnar Fant (1960): Acoustic theory of speech production
  - Gerold Ungeheuer (1962): Elemente einer akustischen Theorie der Vokalartikulation

# Quelle-Filter-Modell der Sprachproduktion



# Quelle-Filter-Modell der Sprachproduktion

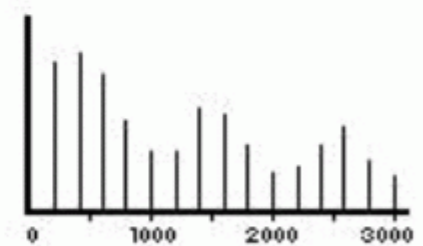
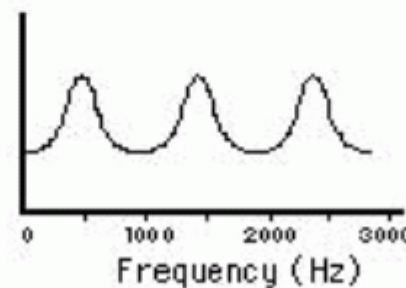
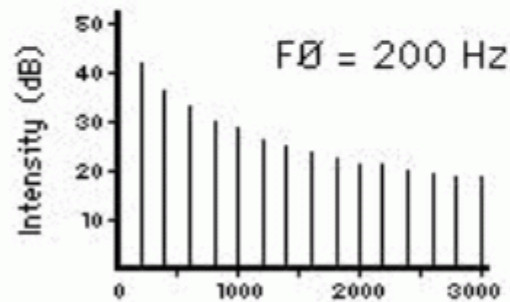
**Glottal Pulses**



**Vocal Tract**



**Speech Signal**

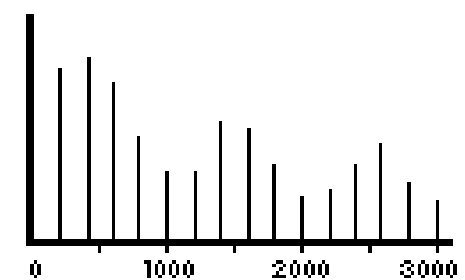
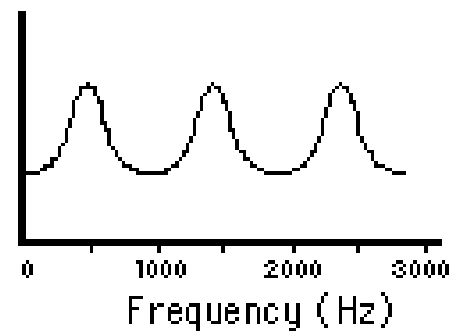
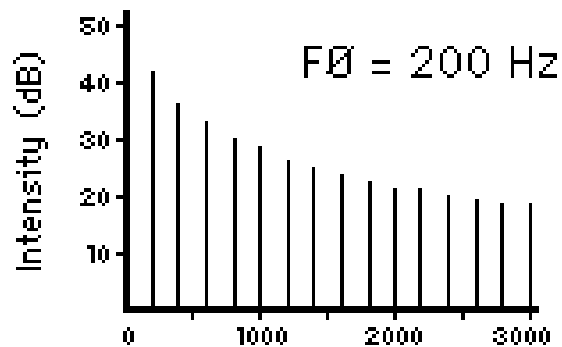
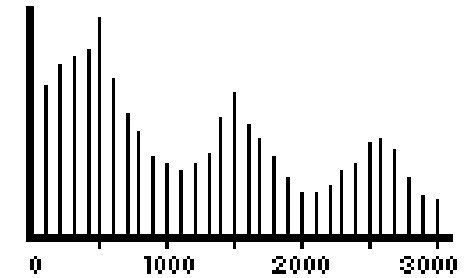
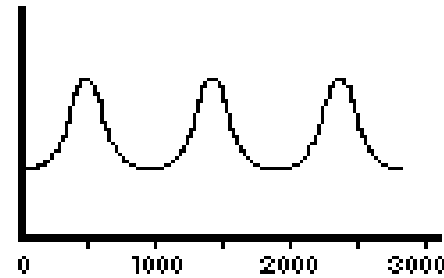
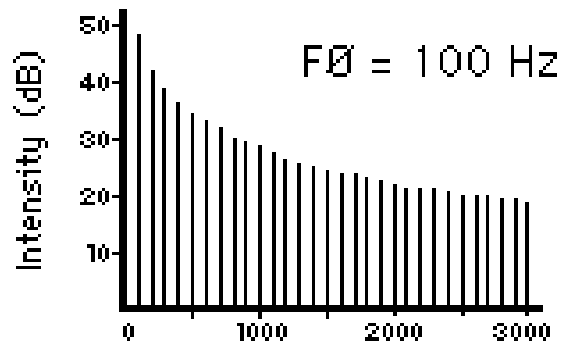


**(a) Source Spectrum**

**(b) Filter Function**

**(c) Output Energy Spectrum**

# Quelle-Filter-Modell der Sprachproduktion



**SOURCE SPECTRUM**

**FILTER FUNCTION**

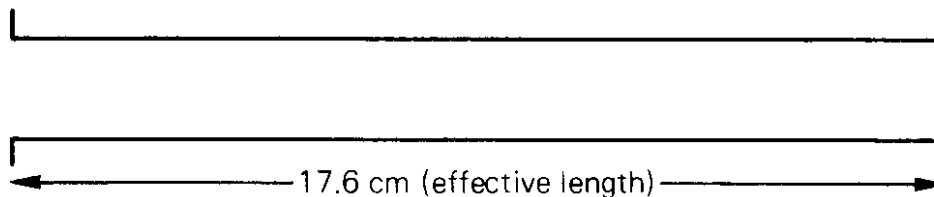
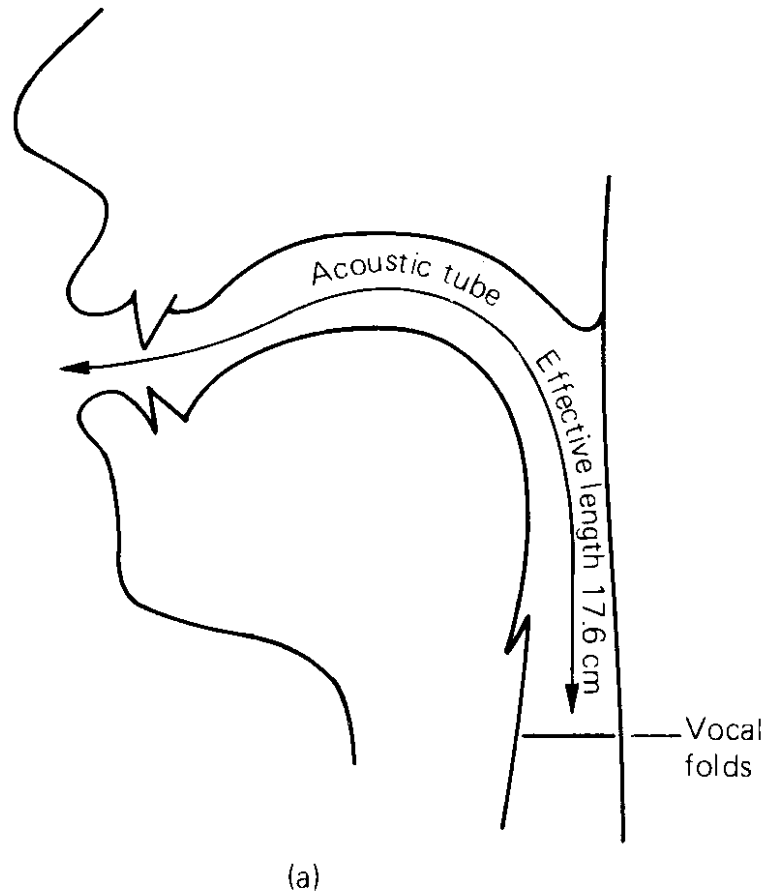
**OUTPUT ENERGY SPECTRUM**

Glottale Anregung

Vokaltrakt: Frequenzantwort

Schallspektrum

# Vokaltrakt: einfaches akustisches Rohrmodell

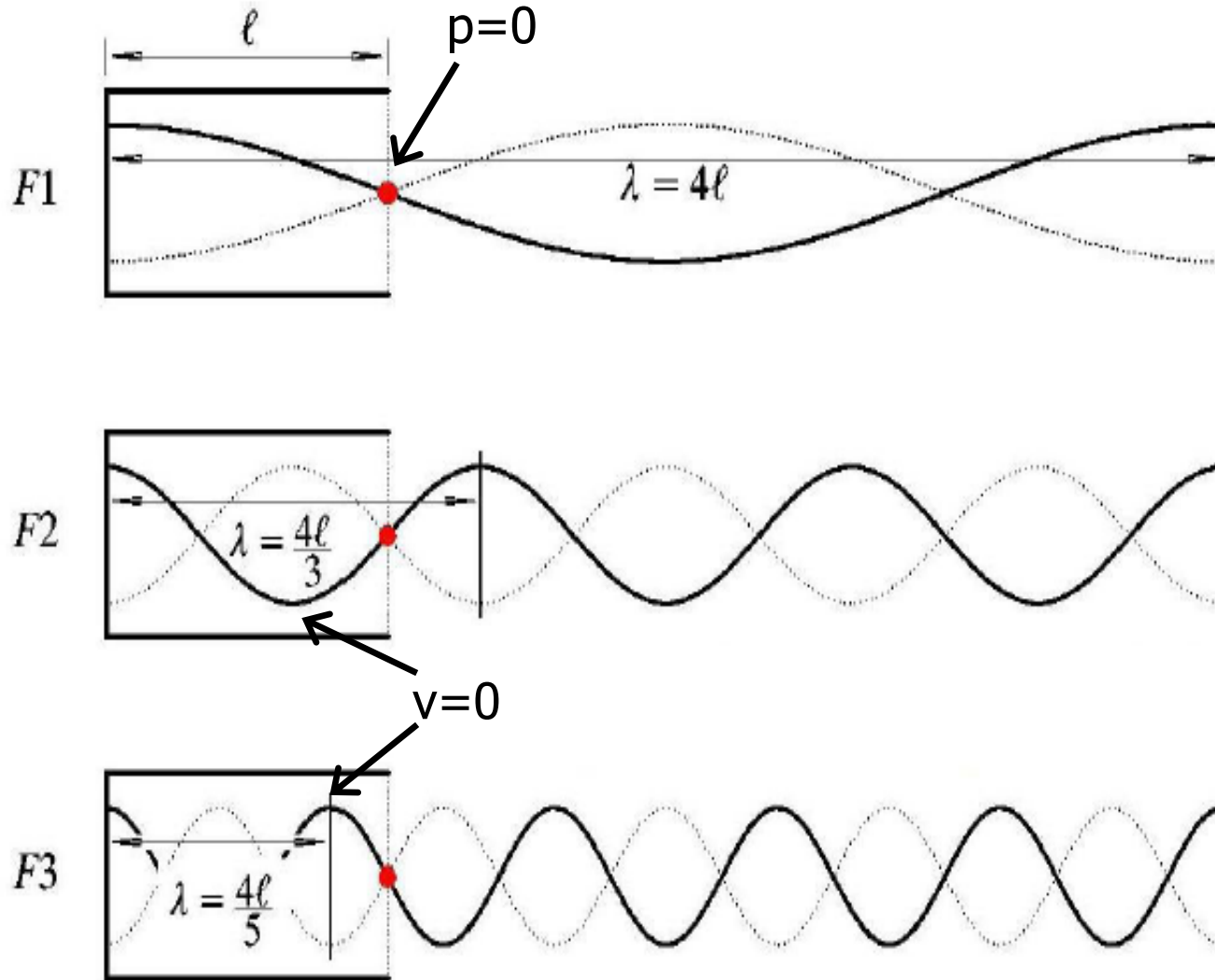


[Clark et al., 2007a, p.241]

# Vokaltrakt: einfaches akustisches Rohrmodell

- Akustische Signale breiten sich im “Ansatzrohr” in Form von Longitudinalwellen aus
- 2 physikalische Größen zur Beschreibung von Wellen:
  - Schalldruck  $p$  : Luftdruckänderungen durch Schall (am Ort der Messung)
  - Schallschnelle  $v$  : Geschwindigkeit, mit der sich Luftteilchen infolge eines Schallereignisses bewegen (nicht: Schallgeschwindigkeit  $c$ )
- (Verlustfreie) Reflexion an schallharter Wand des Rohrs
  - $v = 0$  am Ort der Reflexion
- (Verlustbehaftete) Reflexion an schallweichem Übergang vom Ansatzrohr ins freie akustische Feld (von den Lippen zur Luft)
  - $p = 0$  am Ort der Abstrahlung

# Schalldruckwellen im Vokaltrakt



[Hess, ms.]

# Resonanz- (Formant-) Frequenzen

- Berechnung der Resonanzfrequenzen des neutralen Ansatzrohrs (verlustfrei, konstanter Querschnitt) als Quotient aus Schallgeschwindigkeit ( $c = 340 \text{ m/s}$ ) und Wellenlänge:  $f_i = c / \lambda_i$
- Frequenzen der Resonanzen (Formanten):
  - $F1 = 340 / (4 * 0.17) = 340 / 0.68 = 500 \text{ Hz}$
  - $F2 = 340 / (4/3 * 0.17) = 3 * 340 / (4 * 0.17) = 1500 \text{ Hz}$
  - $F3 = 340 / (4/5 * 0.17) = 5 * 340 / (4 * 0.17) = 2500 \text{ Hz}$
- Diese Verteilung der Resonanzfrequenzen im neutralen Ansatzrohr entspricht der Lage der Formanten des "neutralen" Vokals (schwa [ə])
- Das einfache, querschnittsneutrale Rohrmodell ist inadäquat für die Berechnung der Formanten anderer Vokale (Ungeheuer, 1962)

# Akustische Theorie der Vokalartikulation

## 2.3.1 Ausgangspunkt Webster'sche Horngleichung (nach Ungeheuer, 1962)

Wir gehen nun von der Wellengleichung des Schnellenpotentials  $\Phi$  für die Wellenausbreitung in einem Rohr veränderlichen Querschnittes, der sog. Webster'schen Horngleichung aus

$$\frac{\partial^2 \Phi}{\partial x^2} + \frac{1}{A} \frac{\partial \Phi}{\partial x} \frac{dA}{dx} = \frac{1}{c^2} \frac{\partial^2 \Phi}{\partial t^2} \quad (45)$$

mit den bekannten Randbedingungen:

$$v(t) = 0 \Rightarrow \frac{\partial \Phi}{\partial x} = 0 \quad [\text{Glottis, } x = 0] \quad (46)$$

$$p(t) = 0 \Rightarrow \Phi = 0 \quad [\text{Mundöffnung, } x = l] \quad (47)$$

Mit Hilfe der Trennung der Variablen

$$\Phi(x, t) = \varphi(x) \cdot \psi(t) \quad (48)$$

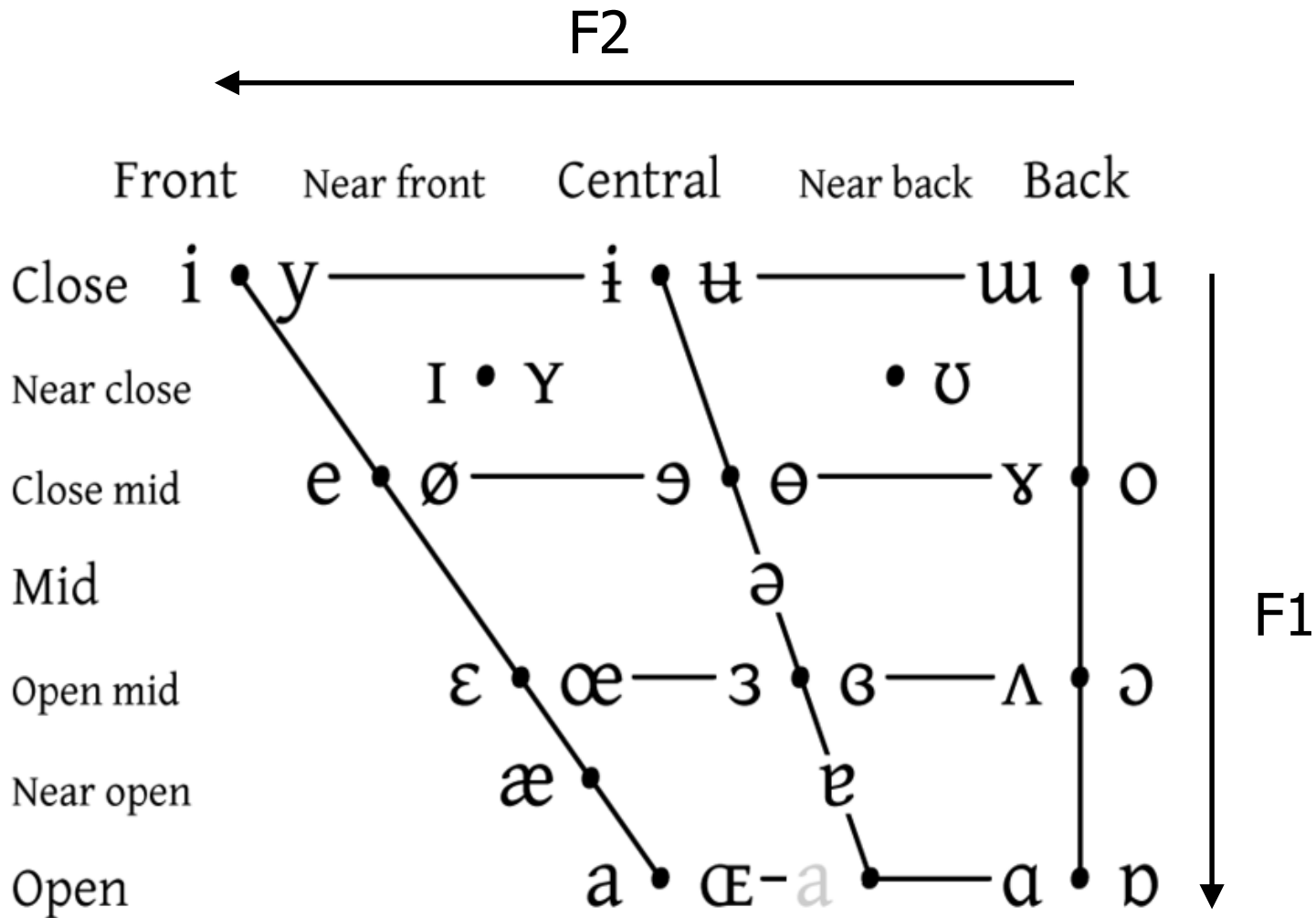
können wir (45) schreiben

$$\frac{1}{\varphi} \left[ \frac{d^2 \varphi}{dx^2} + \frac{1}{A} \frac{d\varphi}{dx} \frac{dA}{dx} \right] = \frac{1}{c^2 \psi} \frac{d^2 \psi}{dt^2} \quad (49)$$

Die linke Hälfte hängt nur von  $x$  ab, die rechte nur von  $t$ . Damit können beide als gleich einer Konstante gesehen werden, die mit  $-\Lambda$  bezeichnet sei:

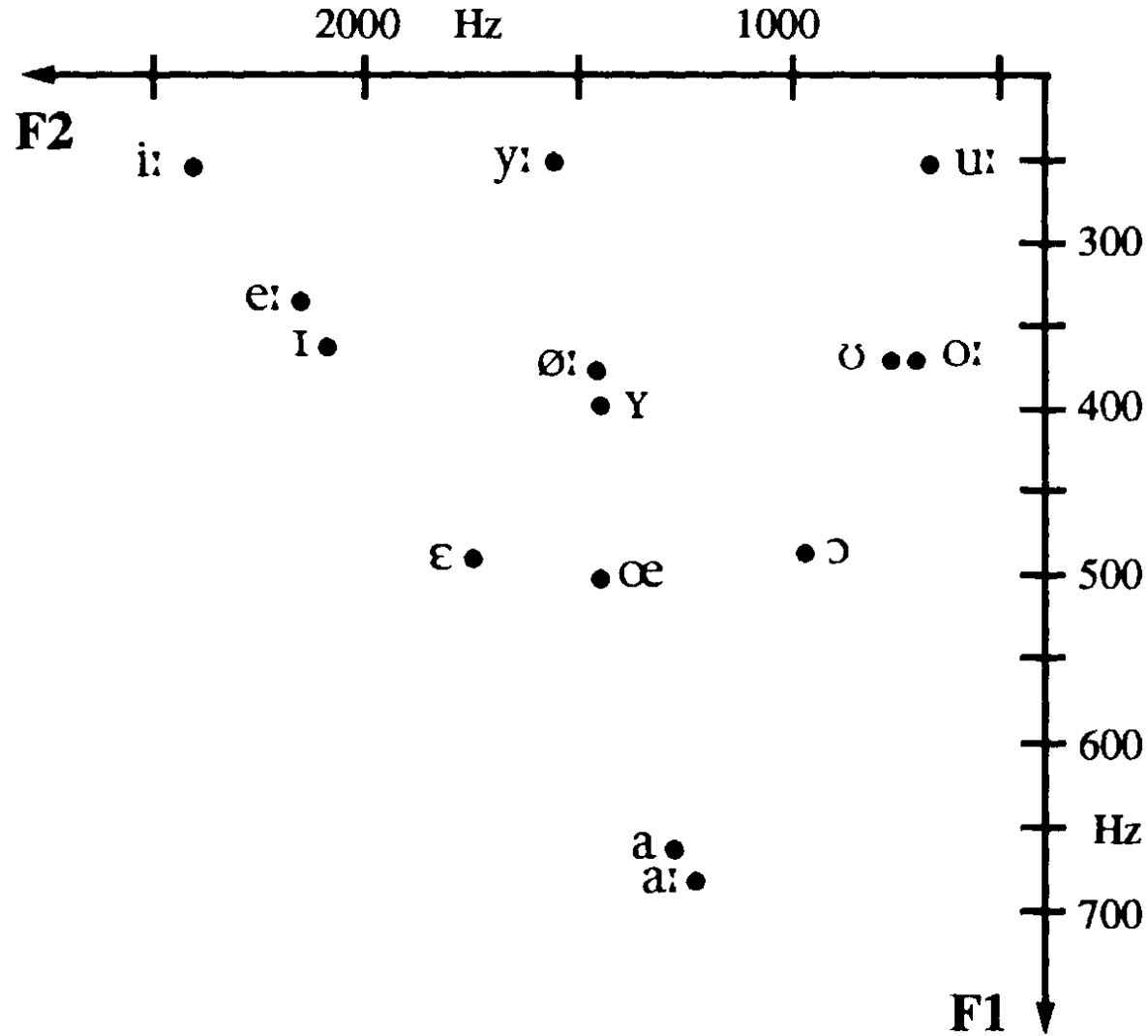
$$\frac{1}{\varphi} \left[ \frac{d^2 \varphi}{dx^2} + \frac{1}{A} \frac{d\varphi}{dx} \frac{dA}{dx} \right] = -\Lambda = \frac{1}{c^2 \psi} \frac{d^2 \psi}{dt^2} \quad (50)$$

# Vokale (IPA)

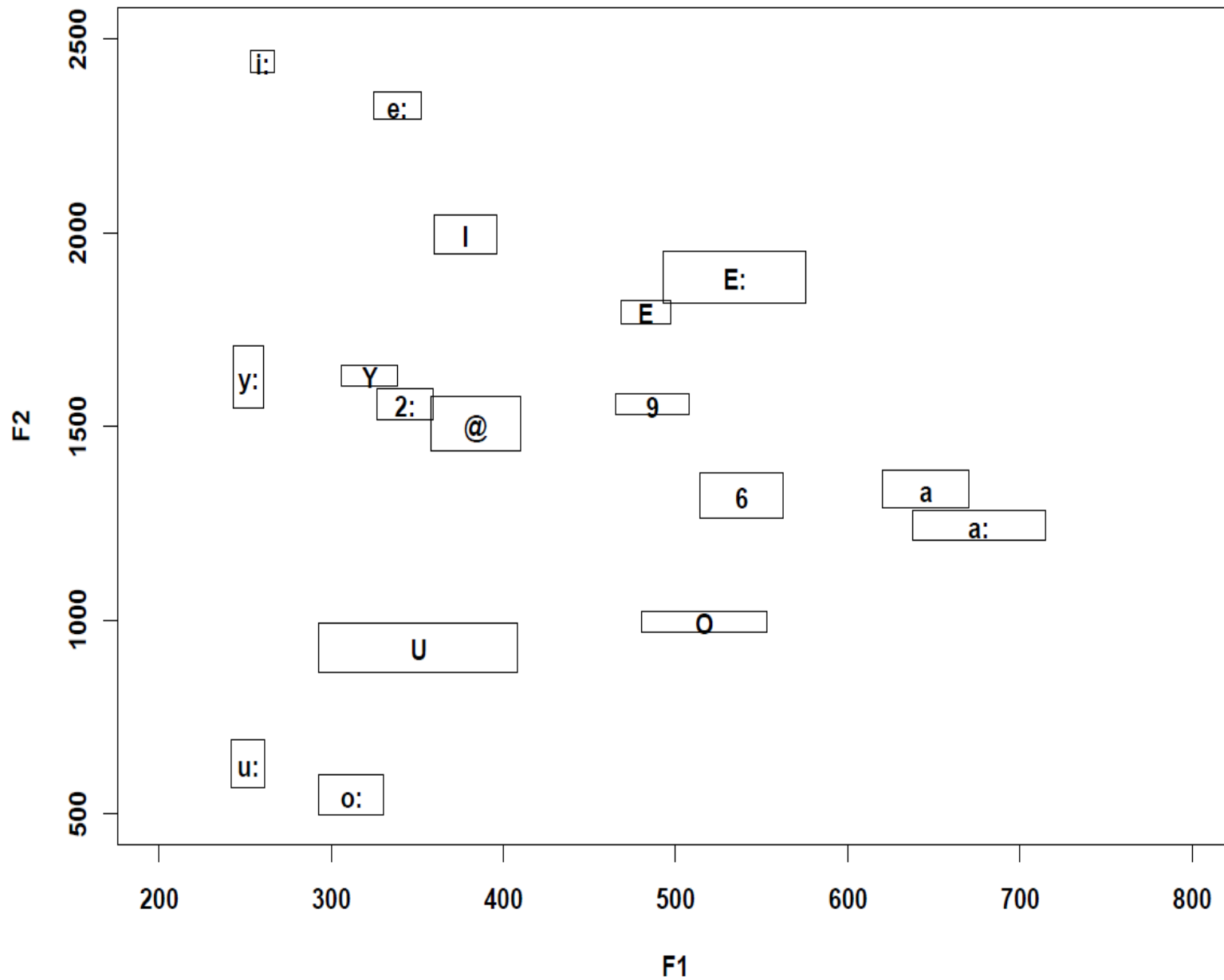


Vowels at right & left of bullets are rounded & unrounded.

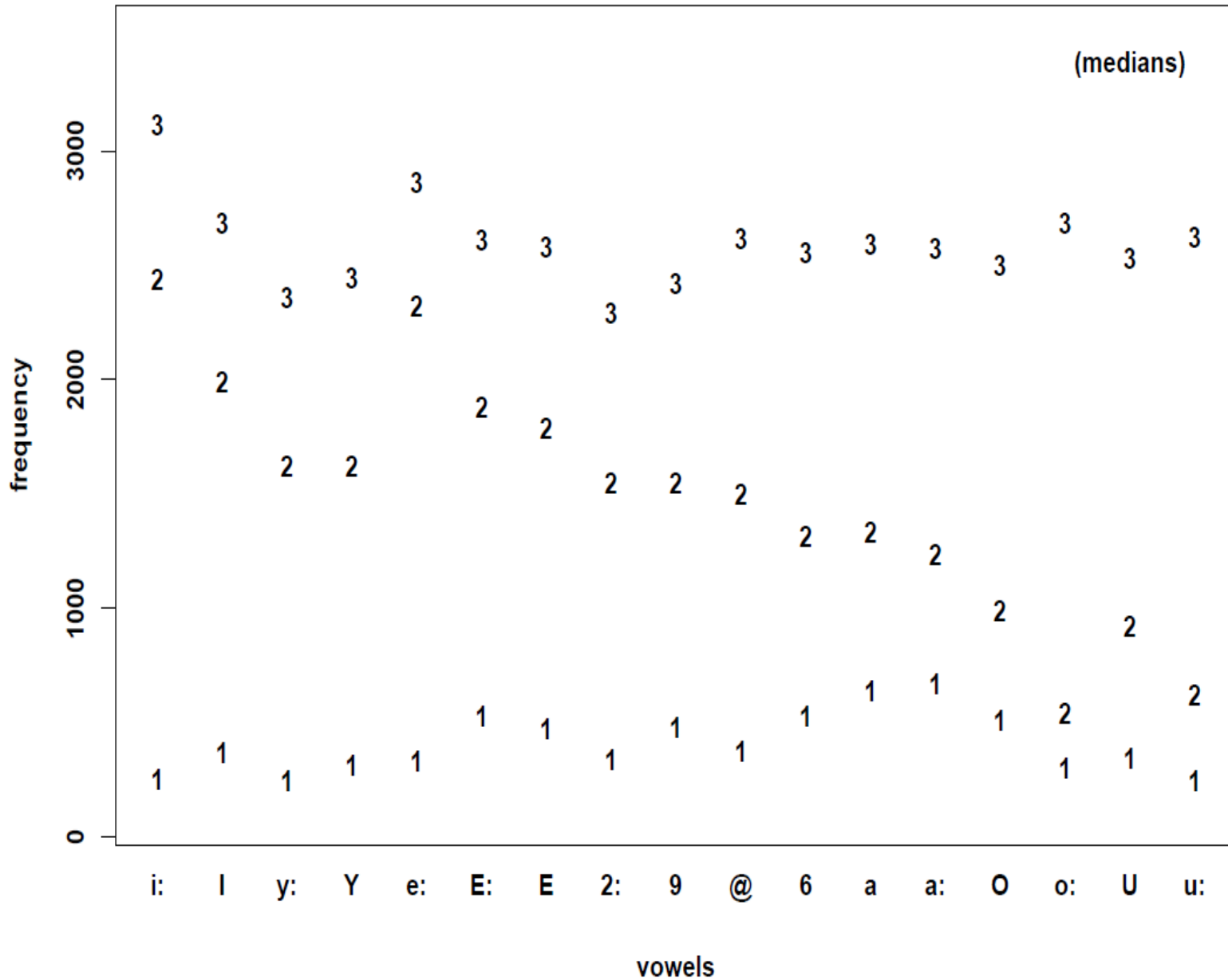
# Vokale (Deutsch [Pompino-Marschall, 1995])



# Vokale (Deutsch [Möbius, 2001])

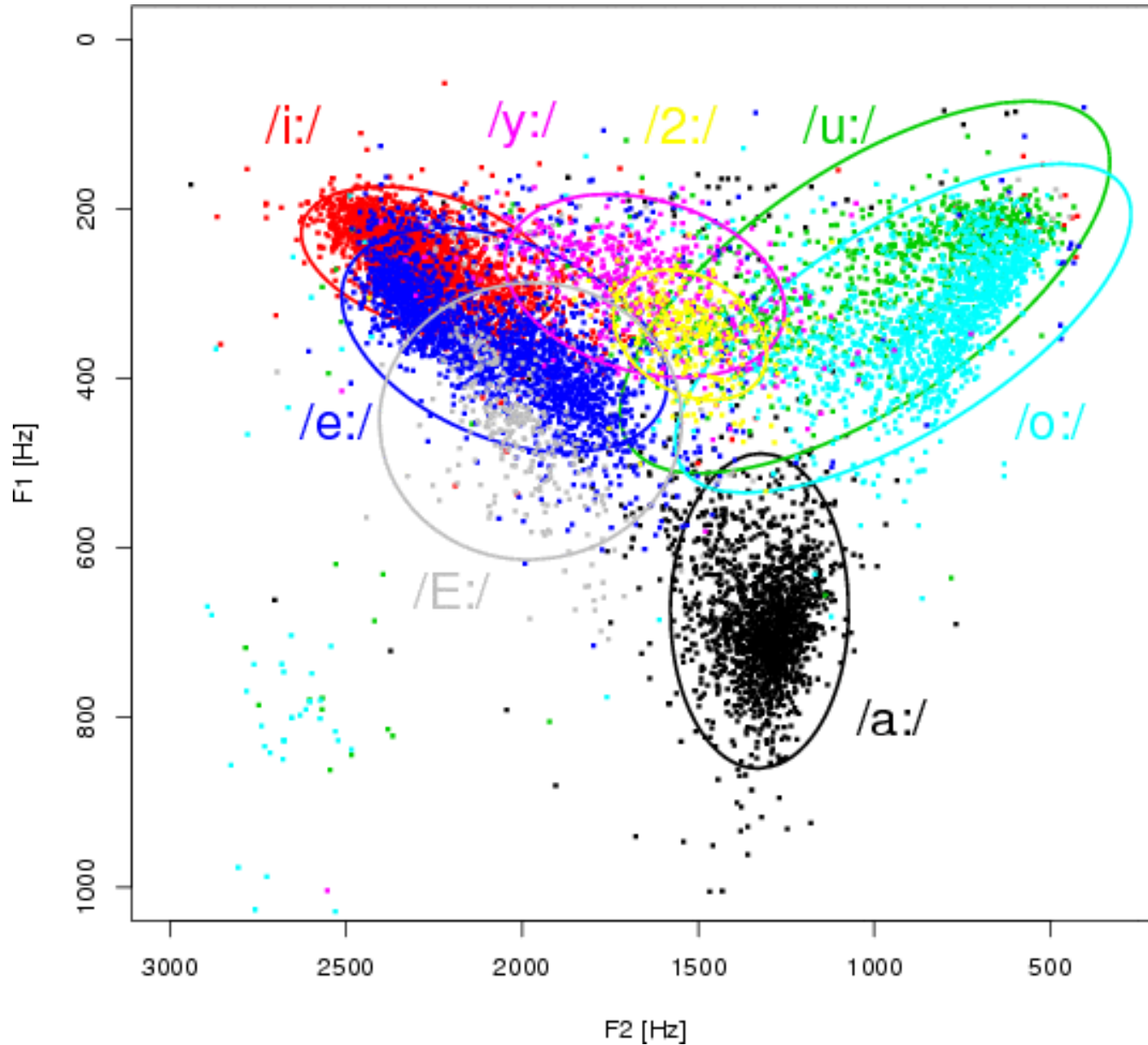


# Vokale (Deutsch, F1/F2/F3 [Möbius, 2001])





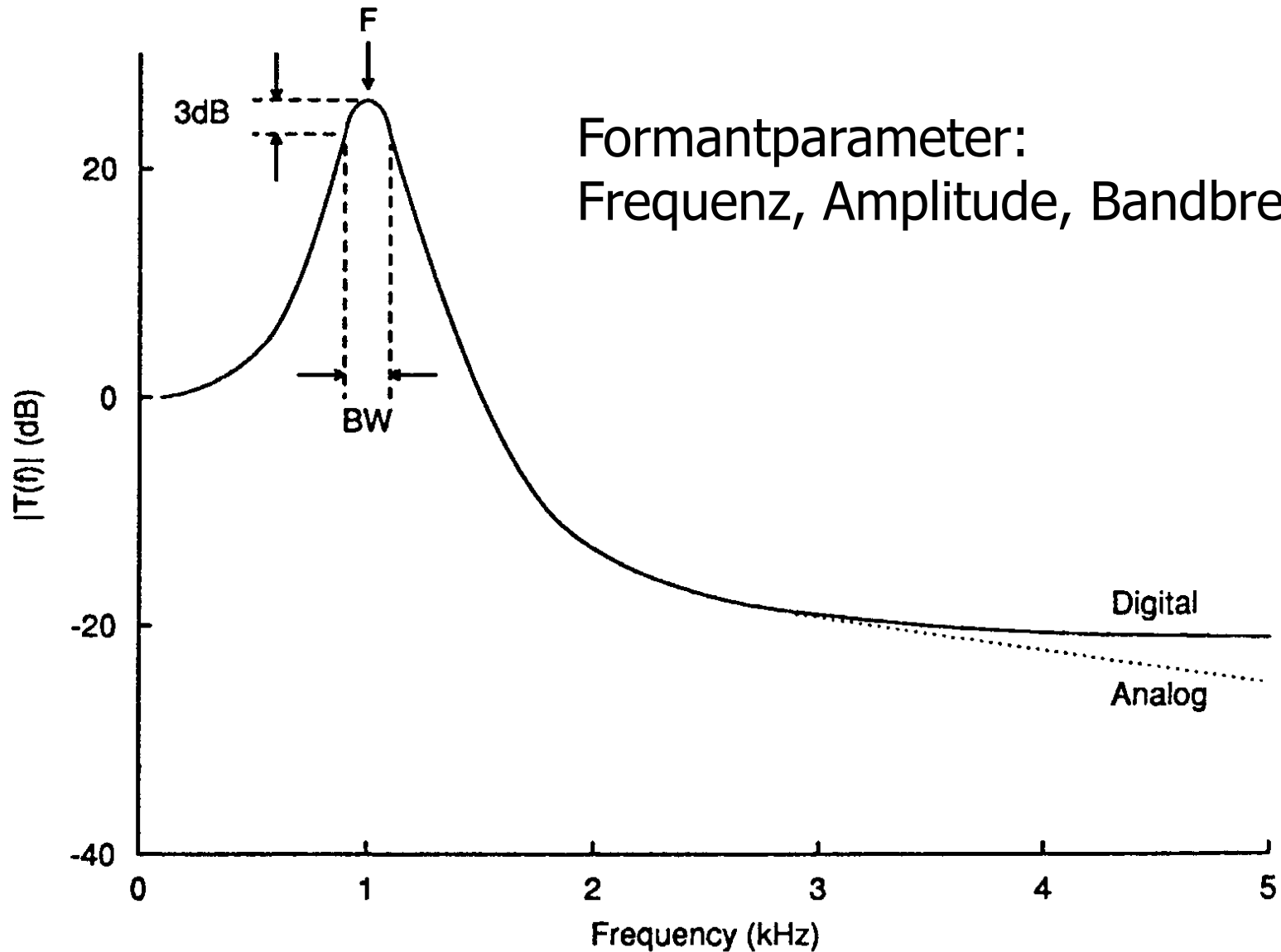
# Vokale (Deutsch [Möbius])



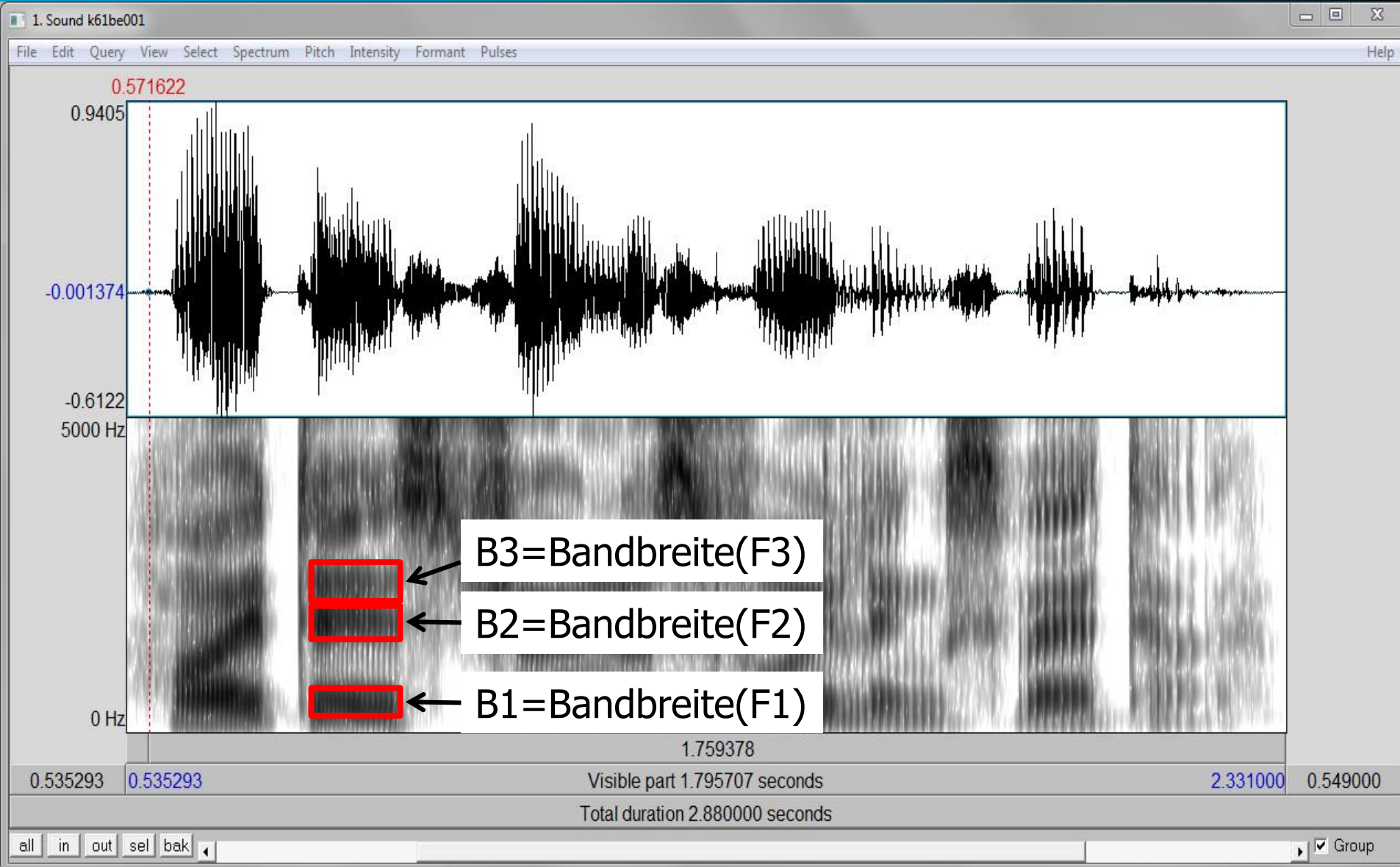
# Vokaltrakt vs. verlustfreies Ansatzrohr

- Im Vokaltrakt entstehen **Verluste** durch
  - Reibung der Luftteilchen untereinander
  - Mitschwingen der Wände des Vokaltrakts
  - weiches Gewebe des Vokaltrakts
  - vor allem: Abstrahlung der Schallenergie ins freie akustische Feld
- verlustbehaftete Schwingungen werden exponentiell gedämpft
- spektrales Äquivalent zur Dämpfung: **Bandbreite**
  - definiert als Frequenzband, innerhalb dessen die Schallenergie auf 50% des Maximalwertes abgeklungen ist
  - entsprechend einem Abfall von 3 dB (oder  $0,707 \cdot \text{Amplitudenwert}$ )
    - Schallenergie wird als Leistung in dB ausgedrückt
    - Schallenergie ist proportional dem Quadrat der Amplitude
    - 50% der Leistung = Energiemaximum minus 3 dB
    - $0,5 \cdot \text{Leistung} = \sqrt{0,5} \cdot \text{Amplitude} = 0,707 \cdot \text{Amplitude}$

# Frequenzantwort (digitaler Resonator)



# Spektrogramm: Formanten



# Vom Spektrum zum Spektrogramm

- Leistungsspektrum: "Schnappschuss" an einem bestimmten Zeitpunkt im Sprachsignal
- Spektrogramm: Zeit als 3. Dimension (neben Frequenz und Amplitude)
  - x-Achse: Zeit [s]
  - y-Achse: Frequenz [Hz]
  - z-Achse: Amplitude [dB] (Graustufen- oder Farbdarstellung)
  - Schmalband-Spektrogramm (z.B. 50 Hz): gute Frequenzauflösung
  - Breitband-Spektrogramm (z.B. 300 Hz): gute Zeitauflösung
  - Analysefensterlänge:
    - kurzes Zeitfenster: gute Zeitauflösung
    - langes Zeitfenster: gute Frequenzauflösung

Danke!

