Foundations of Language Science and Technology

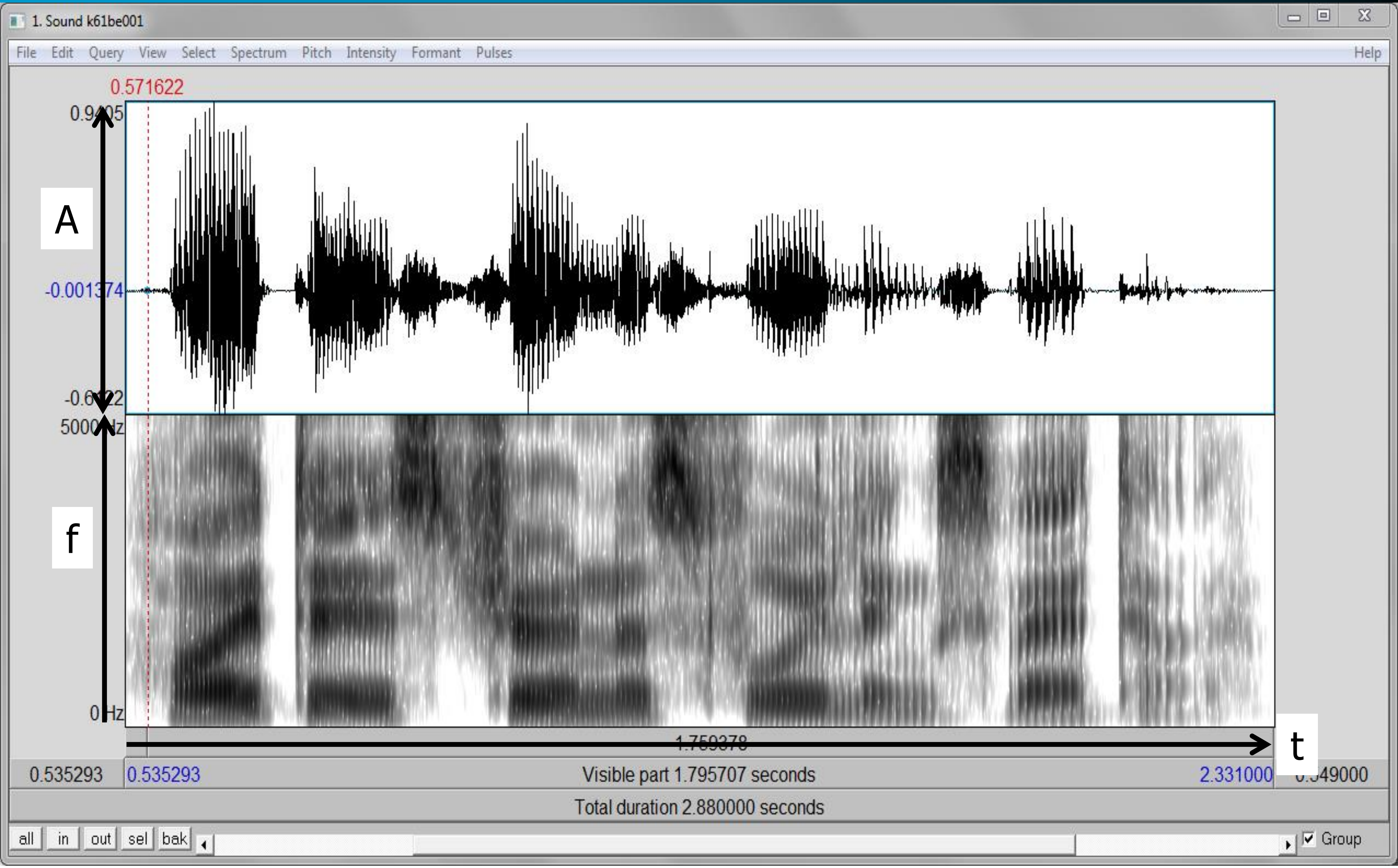
Acoustic Phonetics 1:
Resonances and formants

Jan 19, 2015

Bernd Möbius

FR 4.7, Phonetics
Saarland University

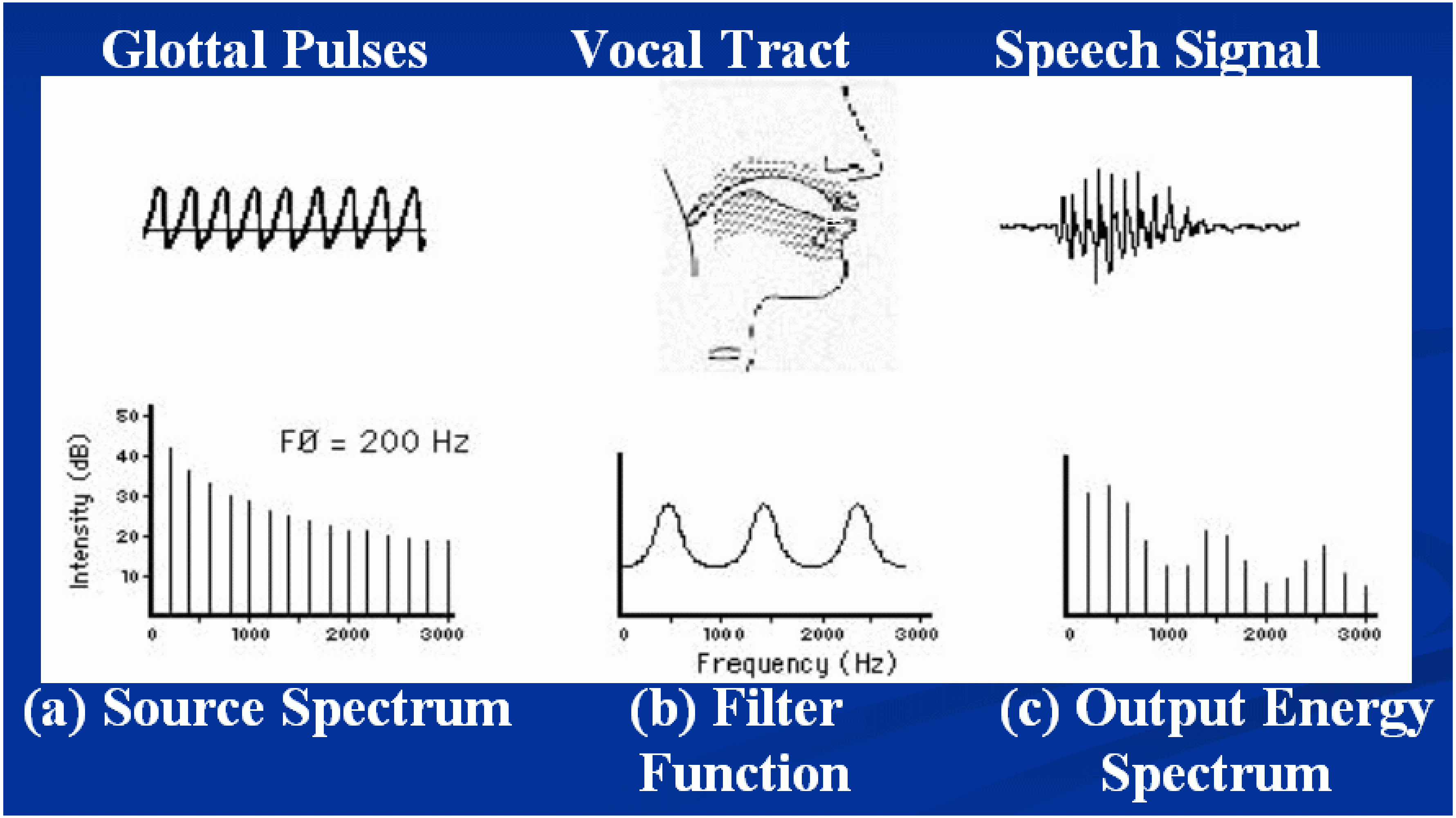
Speech waveforms and spectrograms



Formants

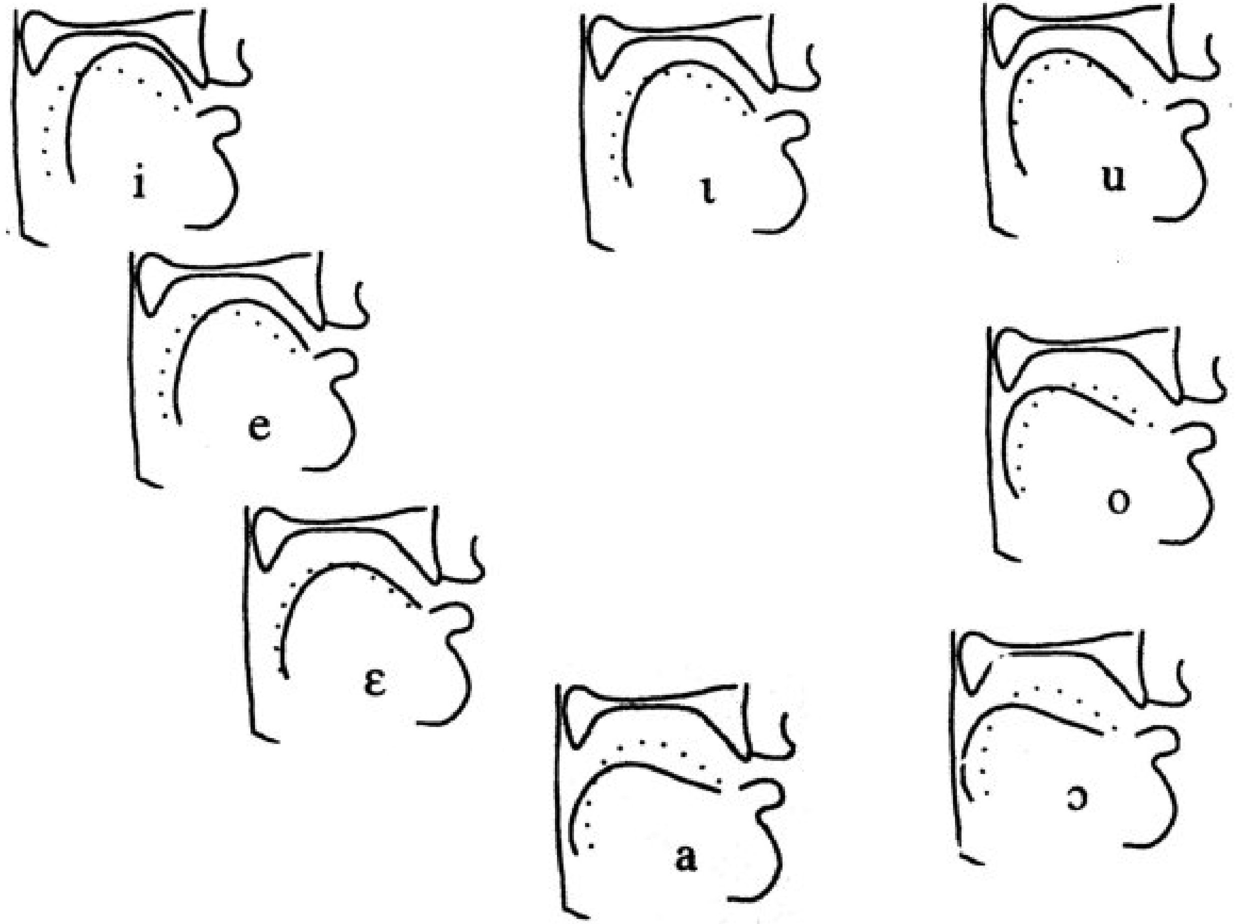
- Spectral peaks, energy maxima: **formants**
- Formants emerge as a consequence of selective reinforcement of certain frequency ranges, corresponding to resonance characteristisc of the vocal tract.
- Distinguishing between *voice source* (periodic, stochastic, transient, mixed excitation) and *sound formation* in the vocal tract motivates the source-and-filter model of speech production.
- References:
  - Gunnar Fant (1960): Acoustic theory of speech production
  - Gerold Ungeheuer (1962): Elemente einer akustischen Theorie der Vokalartikulation

Source-filter model of speech production

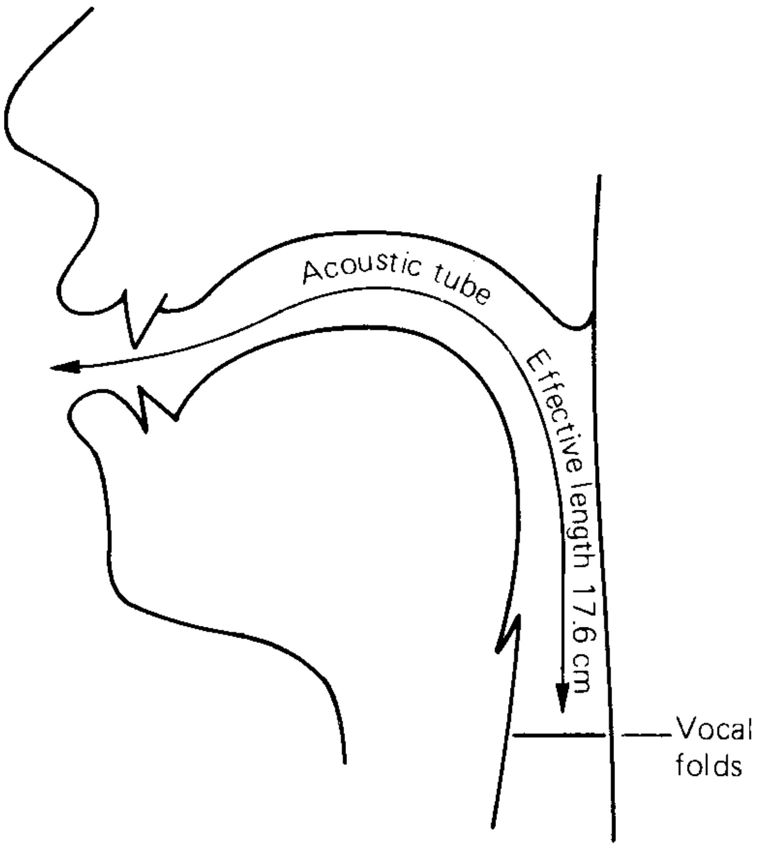


Vocal tract as acoustic filter

- Vocal tract geometry, determined by tongue position, jaw opening, and lip protrusion

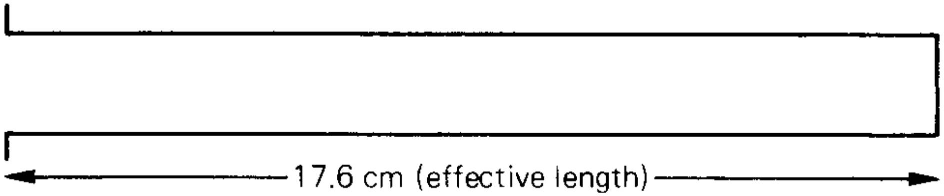


Vocal tract: acoustic tube model



(a)

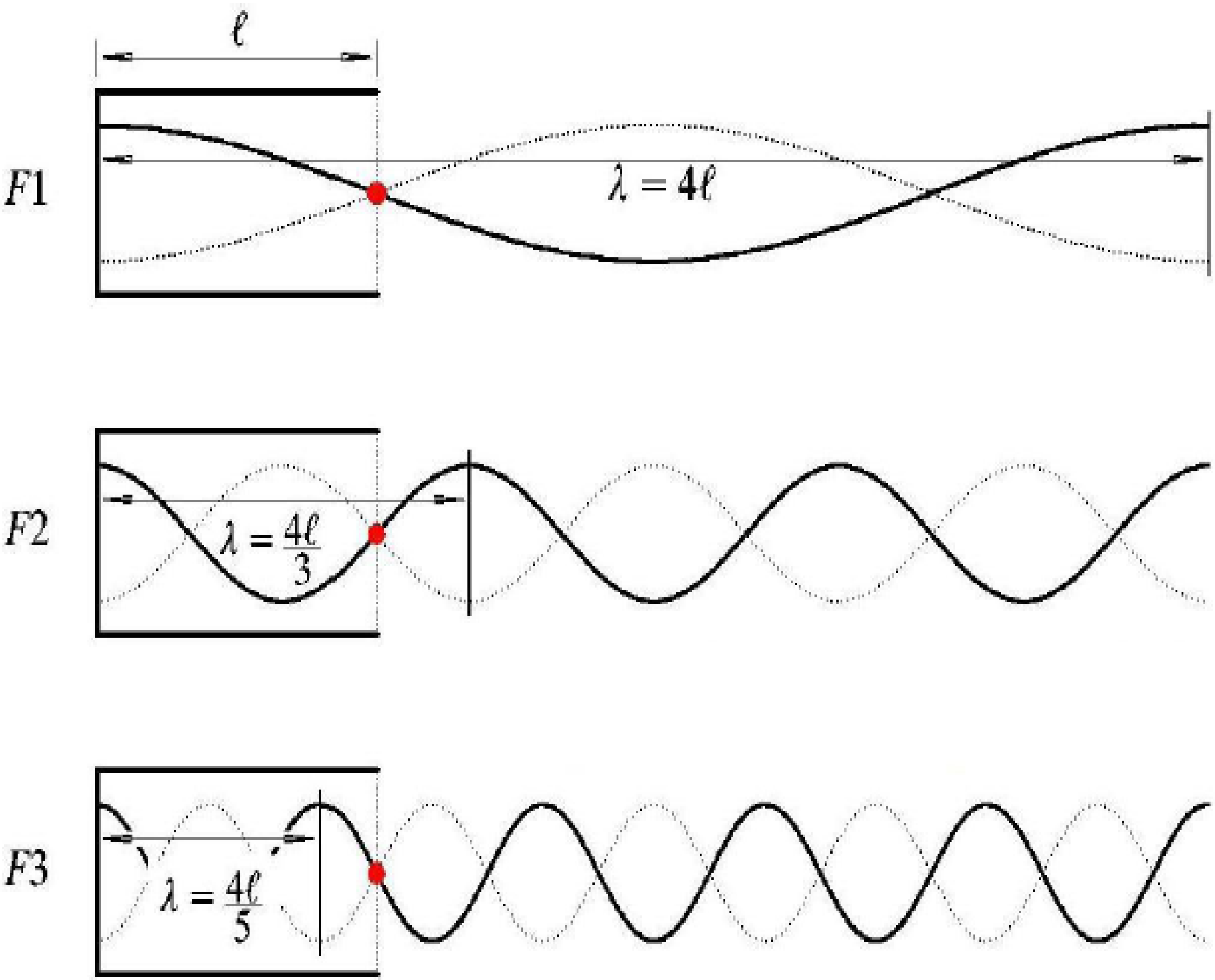
[Clark et al., 2007a, p.241]



Vocal tract: acoustic tube model

- Acoustic signals evolve as longitudinal waves in vocal tract
- 2 physical parameters of acoustic waves
  - sound pressure p : change of air pressure evoked by sound at place of measurement
  - sound velocity v : speed of air particles caused by sound event (note: this is not the speed of sound c !)
- Perfect reflexion at sound-hard (lossless) walls of tube
  - $v = 0$ at place of reflexion
- (Lossy) reflexion at sound-soft transition from vocal tract to free acoustic field (i.e. from lips to air)
  - $p = 0$ at place of radiation

Sound pressure waves in vocal tract



[Hess, ms.]

Computing formant frequencies

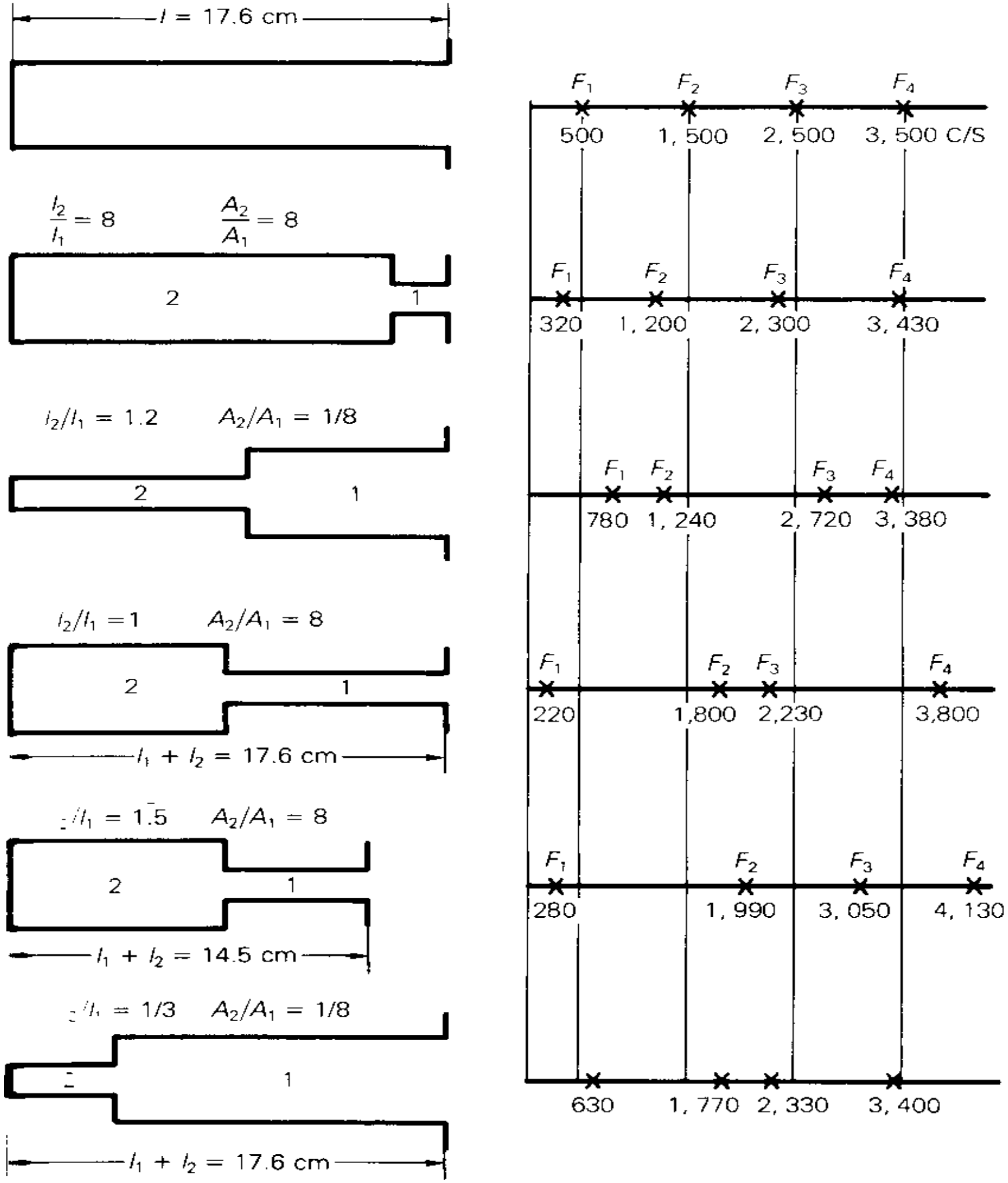
- Resonance frequencies of neutral vocal tract computed as speed of sound divided by wave length: $f_i = c / \lambda_i$
- Frequencies of resonances/formants:

  $F1 = 340 / (4 * 0.17) = 340 / 0.68 = 500$ Hz

  $F2 = 340 / (4/3 * 0.17) = 3 * 340 / (4 * 0.17) = 1500$ Hz

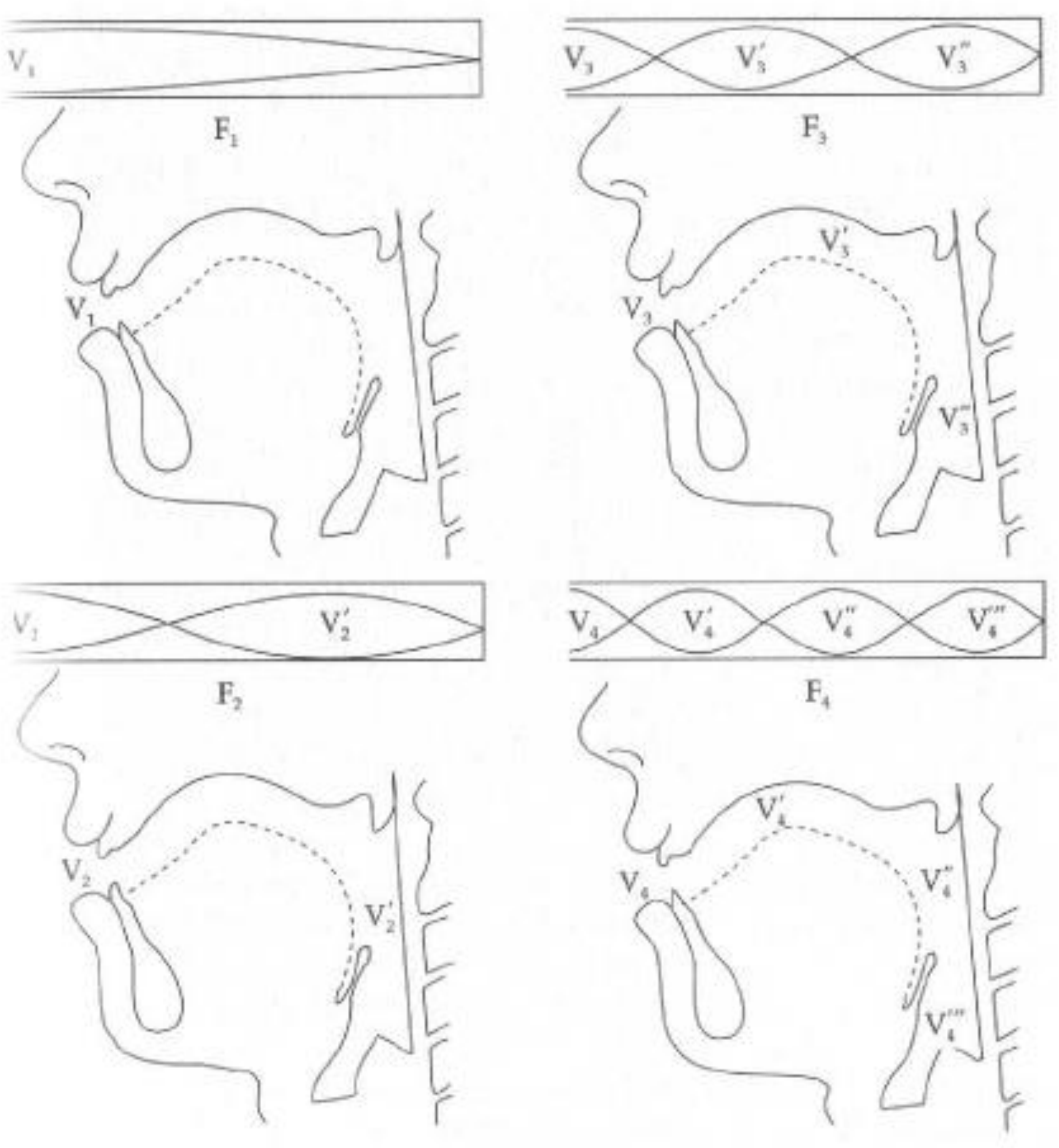
  $F3 = 340 / (4/5 * 0.17) = 5 * 340 / (4 * 0.17) = 2500$ Hz
- Distribution of formant frequencies in neutral vocal tract corresponds to formants of central vowel [ə]
- Simple tube model, with constant area, is inadequate for computing formants of other vowels (cf. acoustic theory of vowel articulation [Ungeheuer 1962])

Tube model with variable area



[Clark et al., 2007a, p.246]

Resonances: standing waves



parameter: v [Johnson, 1997, p.99]

Standing waves: interpretation

- interpretation of the graphical representation of standing waves in idealized vocal tract (neutral configuration, see previous figure):
- first 4 formants displayed (F_1 – F_4)
- in tube model and in vocal tract
- places of maximum sound velocity (sound velocity nodes, V_i)
- places of maximum sound pressure (wave maxima, "antinodes")
- localization of V_i in vocal tract

Dynamic area changes

- resonances of vocal tract with variable area cannot be straightforwardly visualized as in the neutral tube model
  - local area changes affect frequencies of resonances, depending on energy distribution of standing wave in tube along longitudinal axis ("z-axis")
  - e.g., constriction at lip end of tube has same effect as constriction at glottis end: lower resonance frequency
  - acoustic vowel system can be interpreted as representing geometrical changes with respect to neutral tube geometry and resulting changes of resonance frequencies away from neutral values

    → acoustic theory of vowel articulation [Ungeheuer (1962)]

Acoustic theory of vowel articulation

NOT REQUIRED

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

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

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

mit den bekannten Randbedingungen:

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

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

Mit Hilfe der Trennung der Variablen

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

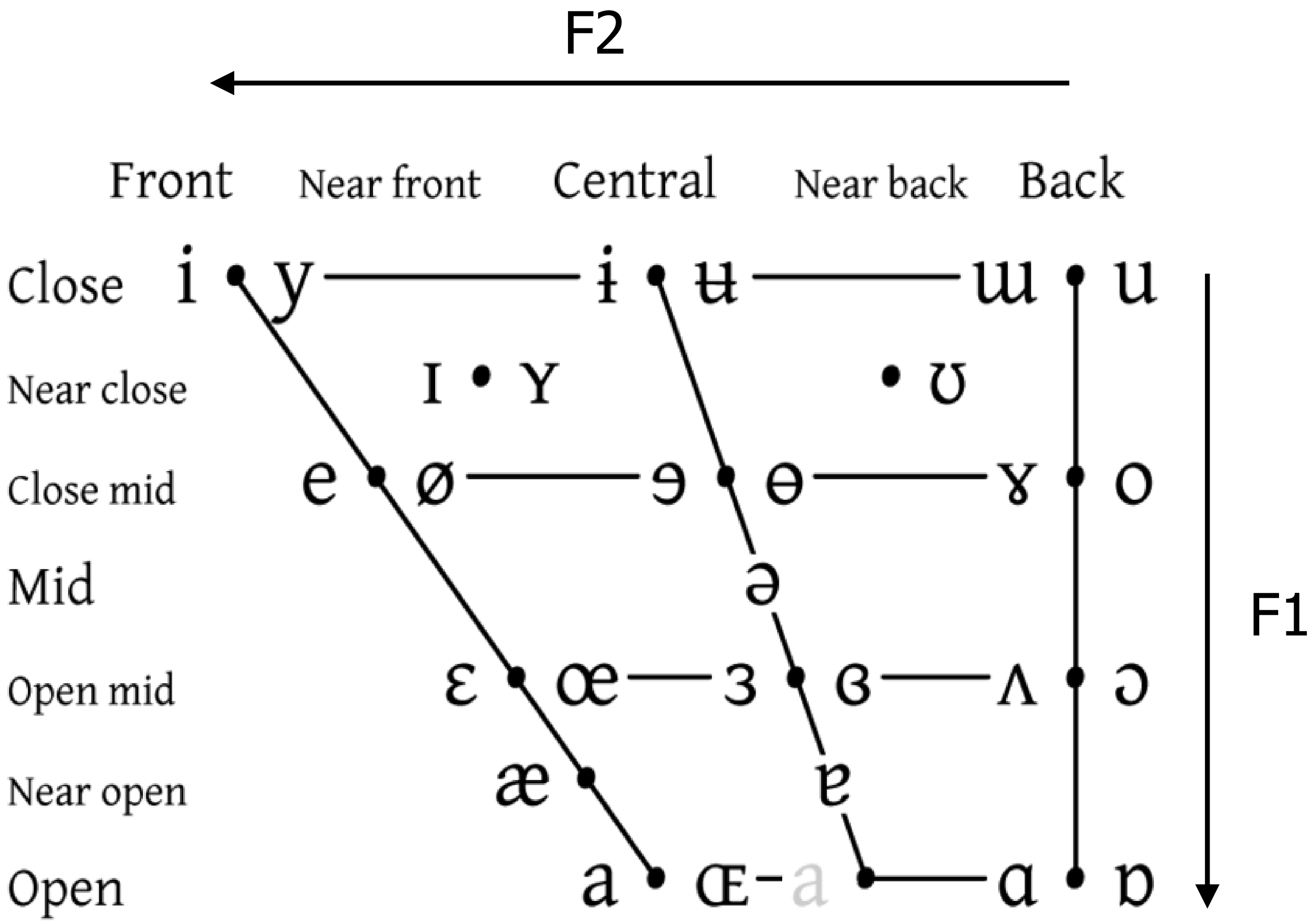
können wir (45) schreiben

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

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:

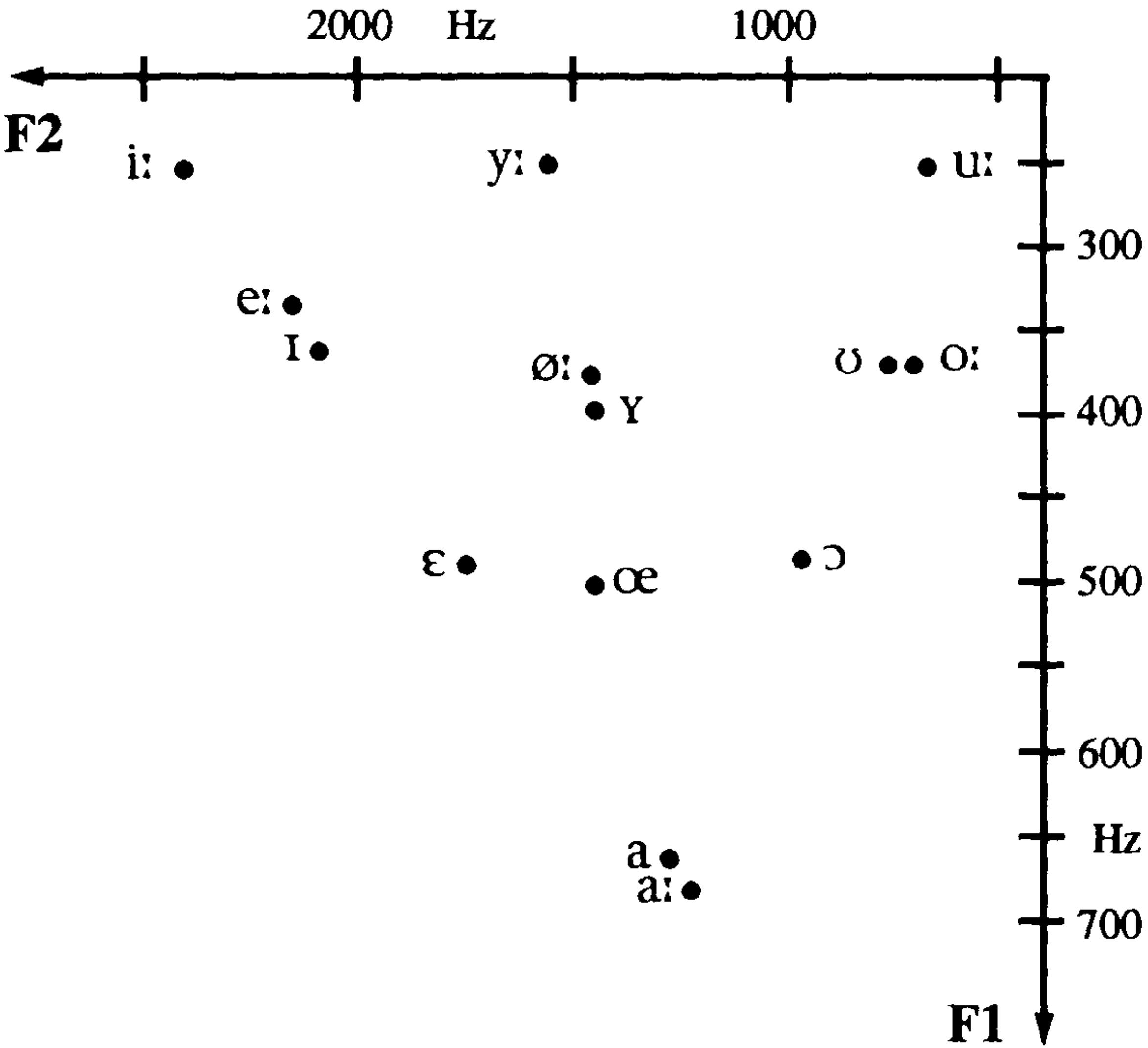
$$\frac{1}{\varphi}\left[\frac{d^2\varphi}{dx^2} + \frac{1}{A}\frac{d\varphi}{dx}\frac{dA}{dx}\right] = -\Lambda = \frac{1}{c^2\psi}\frac{d^2\psi}{dt^2} \qquad (50)$$

Vowels (IPA)

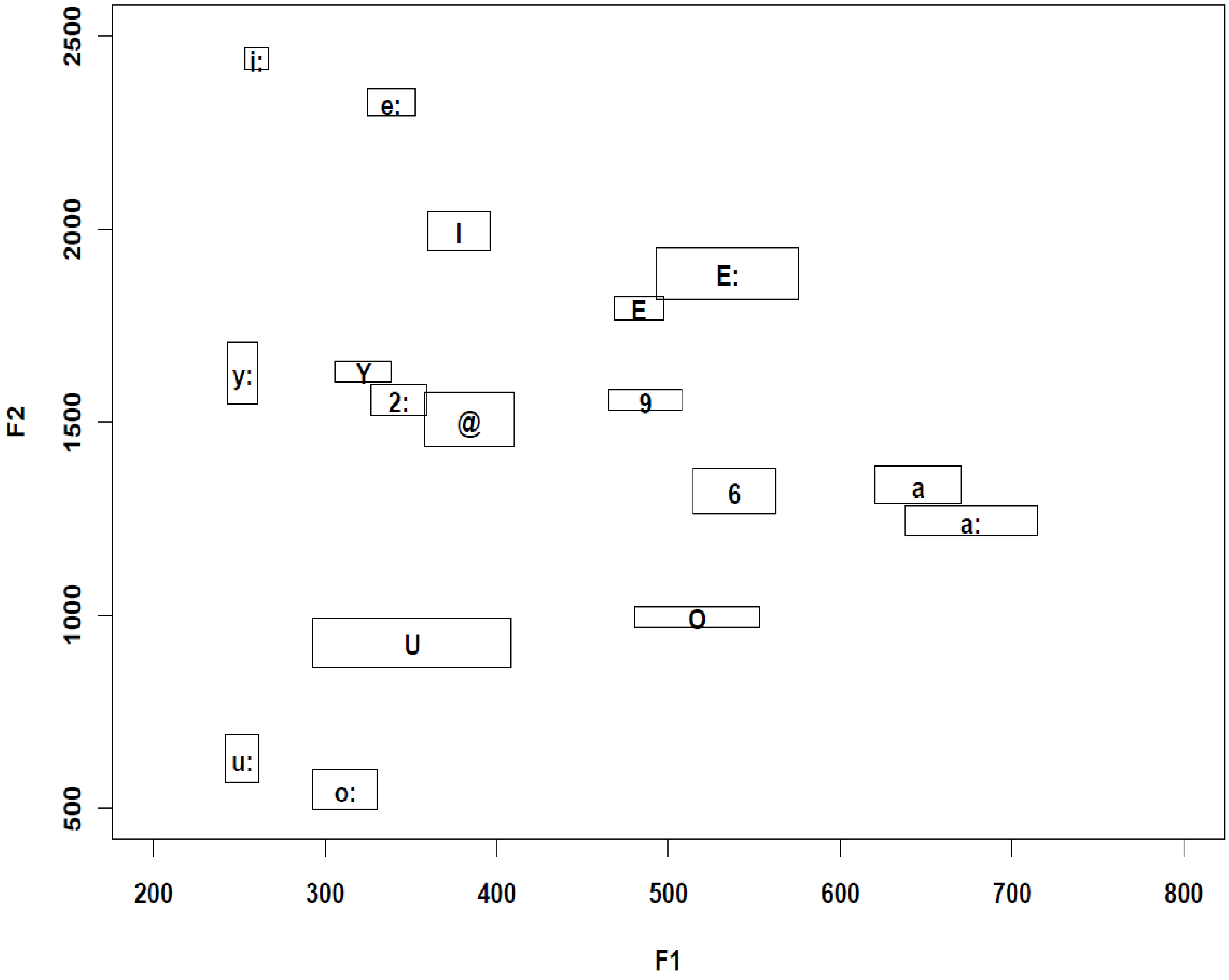


Vowels at right & left of bullets are rounded & unrounded.

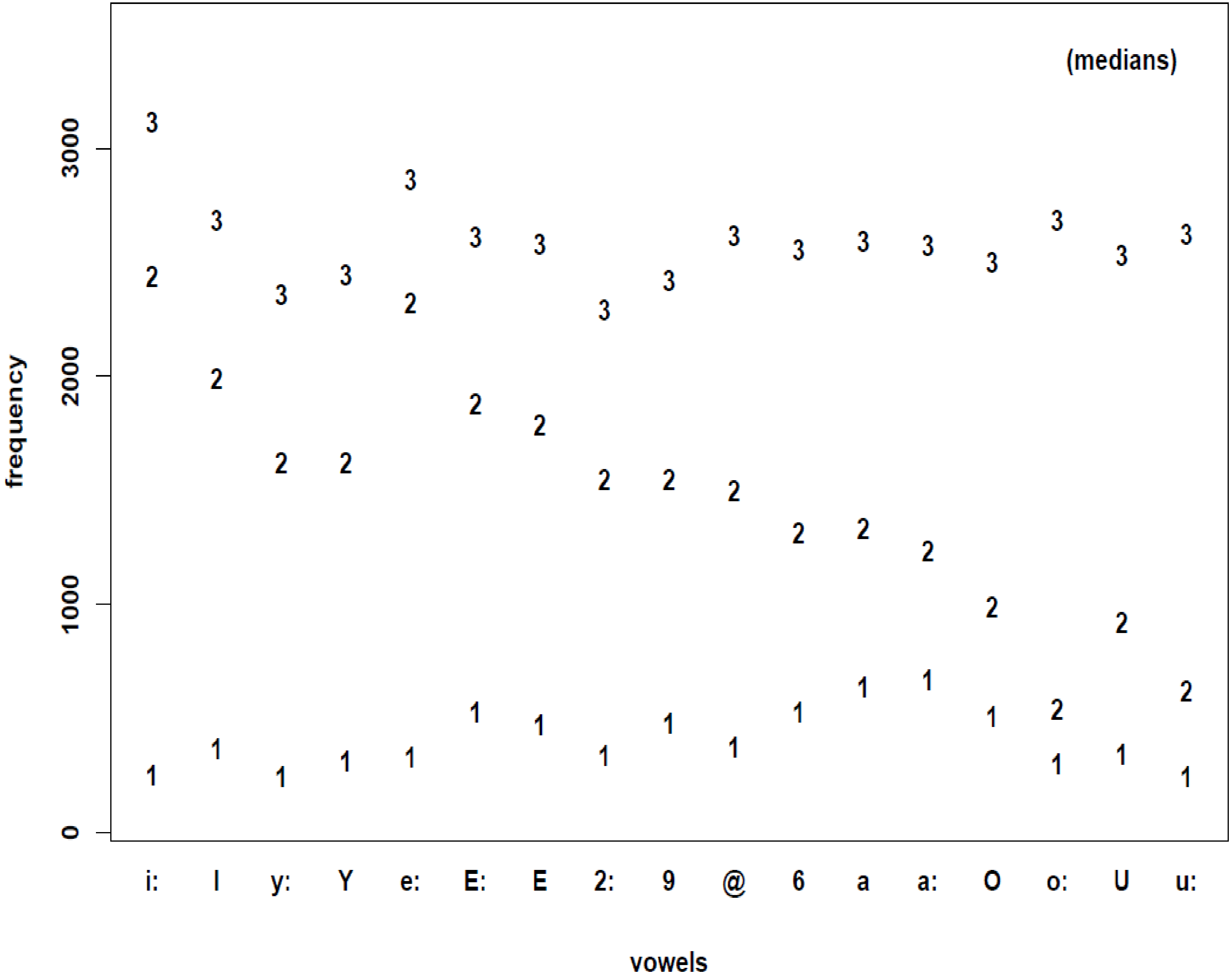
Vowels (German [Pompino-Marschall, 1995])



Vowels (German [Möbius, 2001])

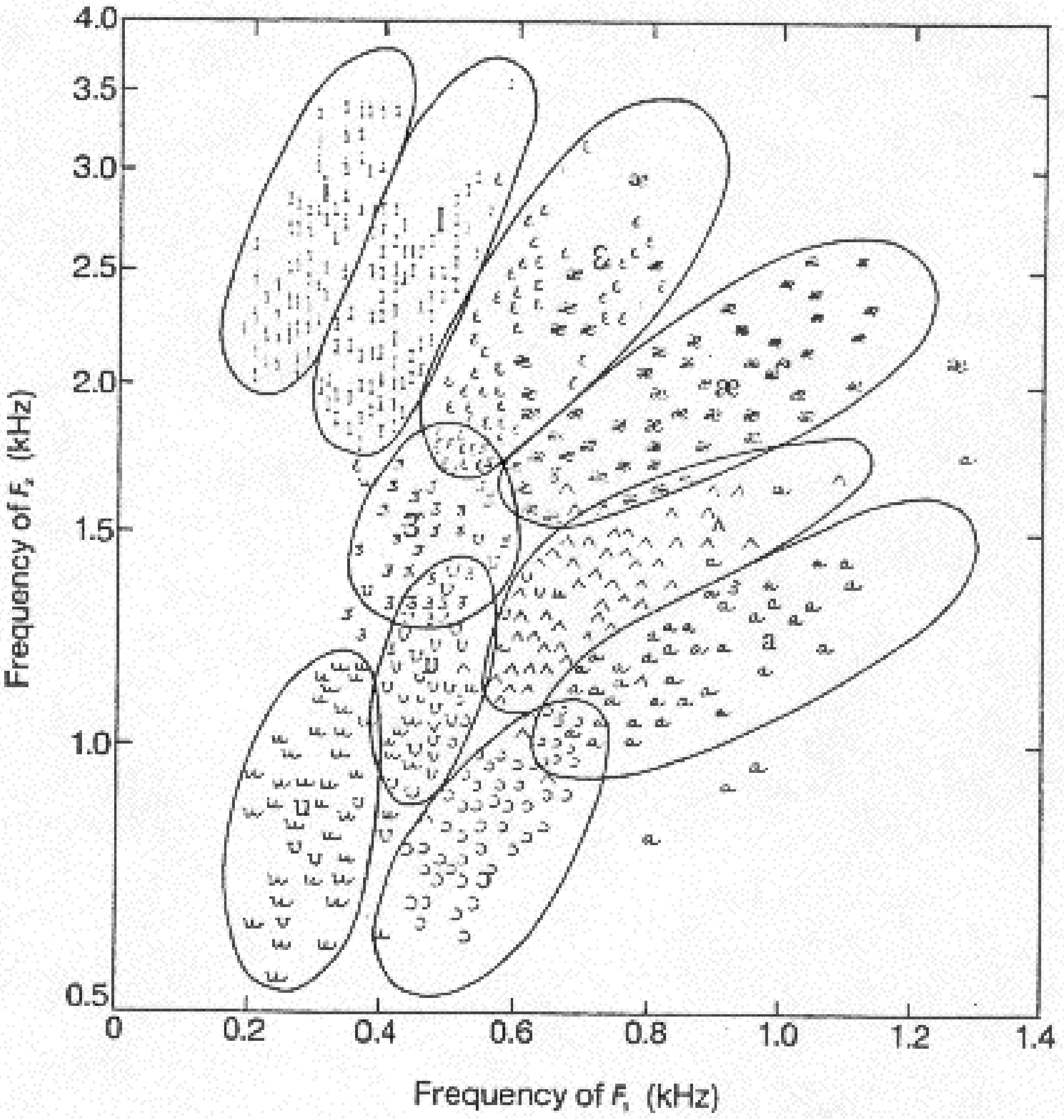


Vowels (German, F1/F2/F3 [Möbius, 2001])

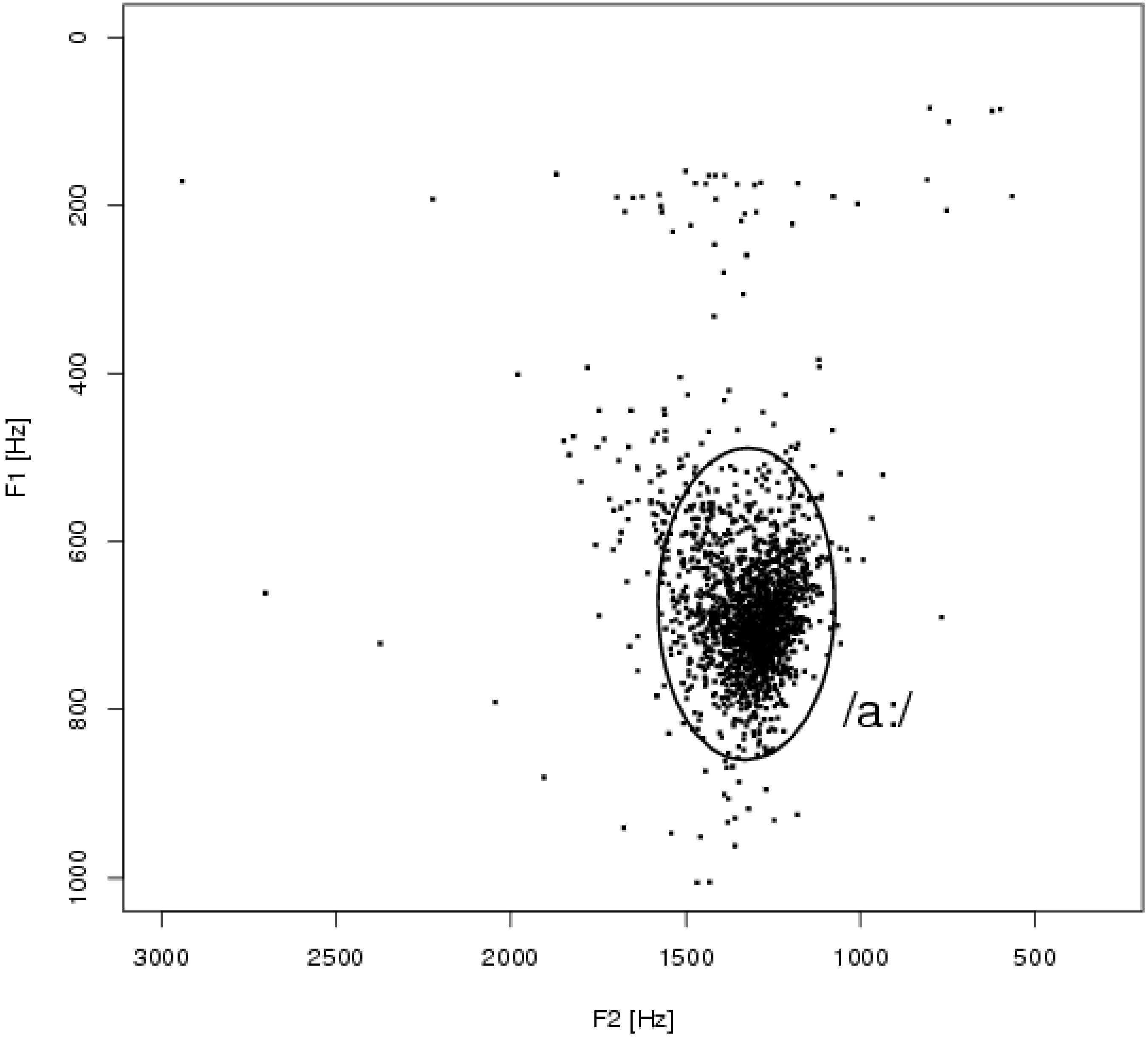


UNIVERSITÄT DES SAARLANDES

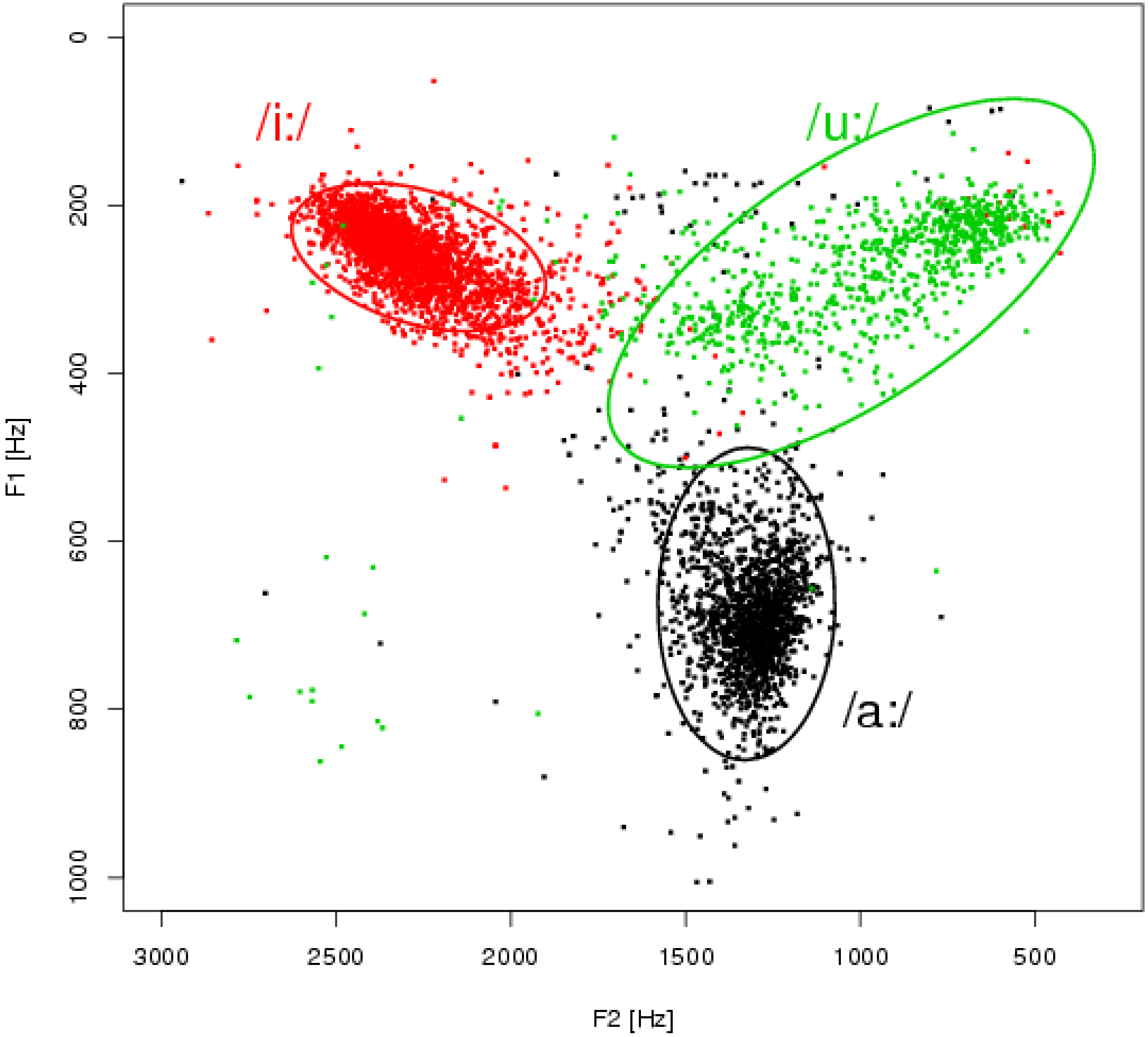
Vowels (Am. English [Peterson and Barney, 1952])



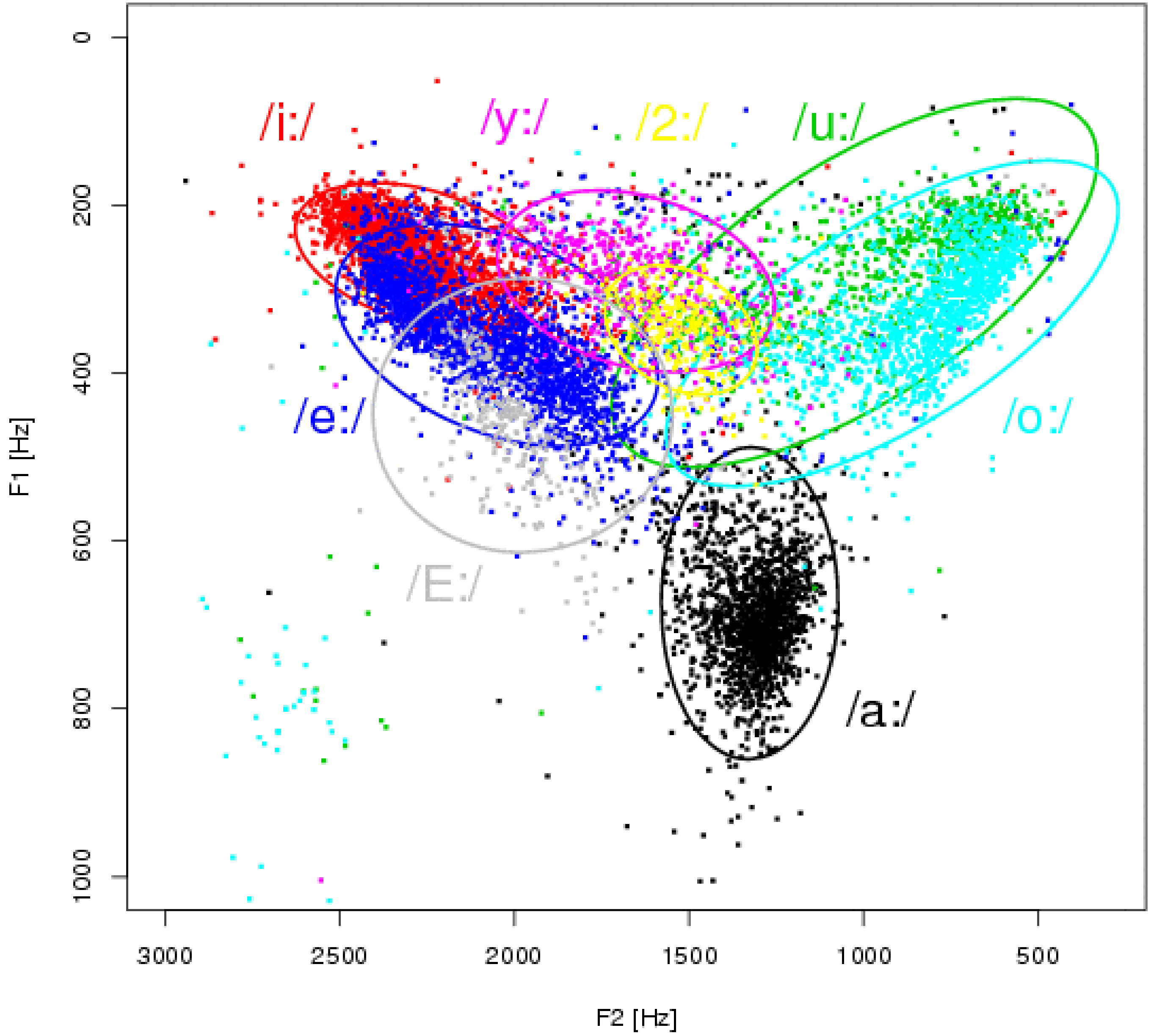
Vowels (German [Möbius])



Vowels (German [Möbius])



Vowels (German [Möbius])

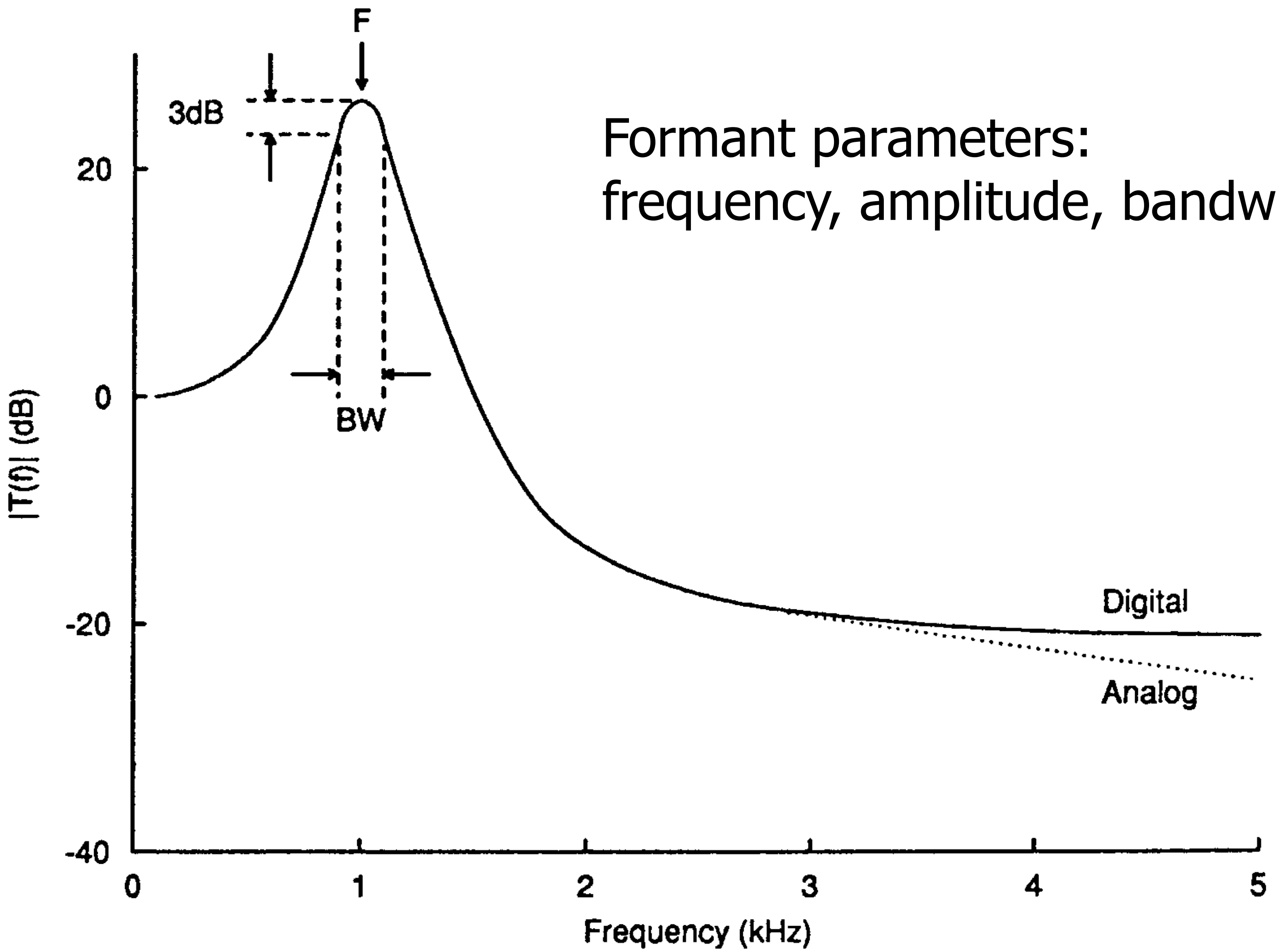


UNIVERSITÄT
DES
SAARLANDES

Vocal tract vs. lossless tube

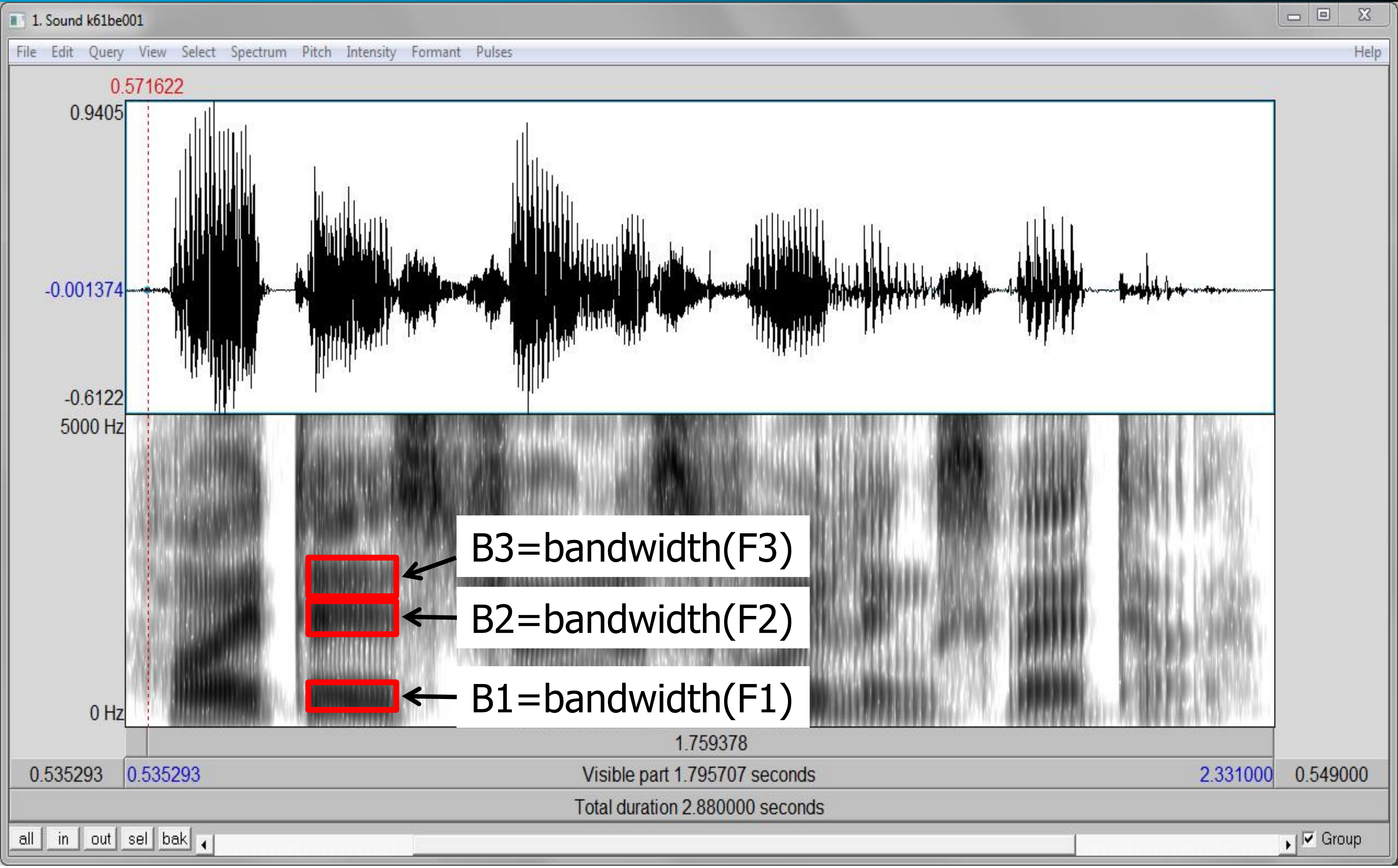
- losses in the vocal tract caused by
  - friction between air particles
  - vibration of vocal tract walls
  - viscosity of vocal tract tissue
  - radiation of sound energy into free acoustic field
- lossy vibrations are damped exponentially
- spectral equivalent of damping: **bandwidth**
  - defined as frequency range comprising 50% of power
  - corresponding to decrease of amplitude by 3 dB (or 0.707*A)
    - sound energy expressed in [dB]
    - sound energy is proportional to square of amplitude
    - 50% of power = energy maximum minus 3 dB
    - 0.5 * power = $\sqrt{0.5}$ * amplitude = 0.707 * amplitude

Resonance response



Formant parameters:
frequency, amplitude, bandwidth

Speech waveforms and spectrograms



Visible part 1.795707 seconds

Total duration 2.880000 seconds

B3=bandwidth(F3)

B2=bandwidth(F2)

B1=bandwidth(F1)

UNIVERSITÄT DES SAARLANDES

Thanks!