

Sprachverarbeitung I / 1    HS 2016

# Einführung — Beschreibung von Sprache

Buch: Kapitel 1

Beat Pfister



# Vorlesung Sprachverarbeitung

Dieses Buch wird für die Vorlesungen  
SPV I und SPV II gebraucht.

<http://www.springer.com/de/book/9783662528372>



# Vorlesung SPV I: Stoffübersicht

Grundsätzliches zur Frage: Was ist Sprache bzw. Lautsprache?	1 DL
Übersicht über die Sprachverarbeitung und Darstellung und Eigenschaften von Sprachsignalen	1 DL
Analyse von Sprachsignalen	3 DL
Einführung in die Sprachsynthese	4 DL
Einführung in die Spracherkennung	4 DL

## Programm heute:

(Buch: Kapitel 1)

Vorlesung: Was ist (lautliche) Sprache?

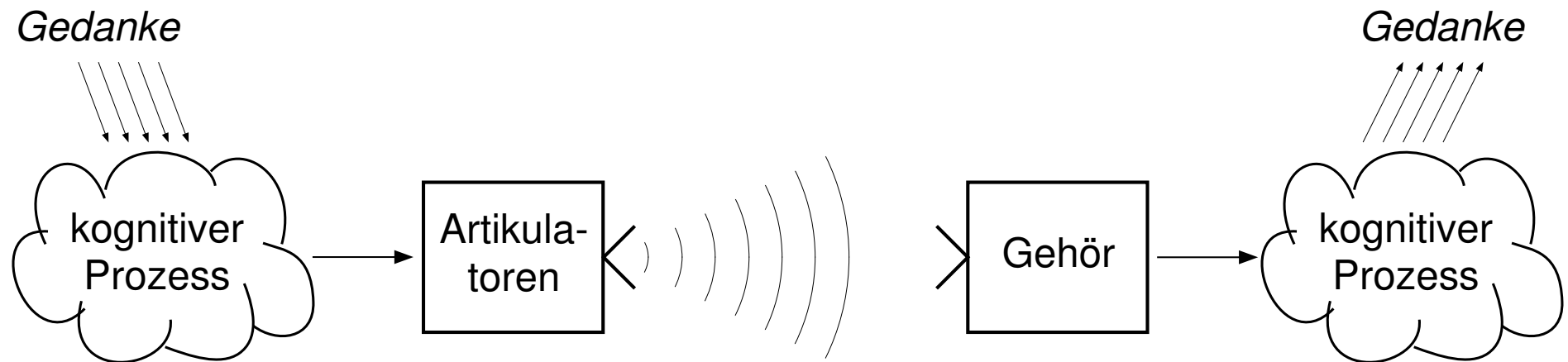
- lautsprachliche Kommunikation
- Beschreibung von Sprache
- Sprachproduktion beim Menschen
- Gehör und Sprachwahrnehmung

Übung: ★ Einführung in Matlab & Audiosignale

# Lautsprachliche Kommunikation zwischen Menschen

**sprechende Person**

**zuhörende Person**



# Lautsprachliche Kommunikation zwischen Mensch und Maschine

Maschine als Sprecher: **Sprachsynthese** Text → Lautsprache

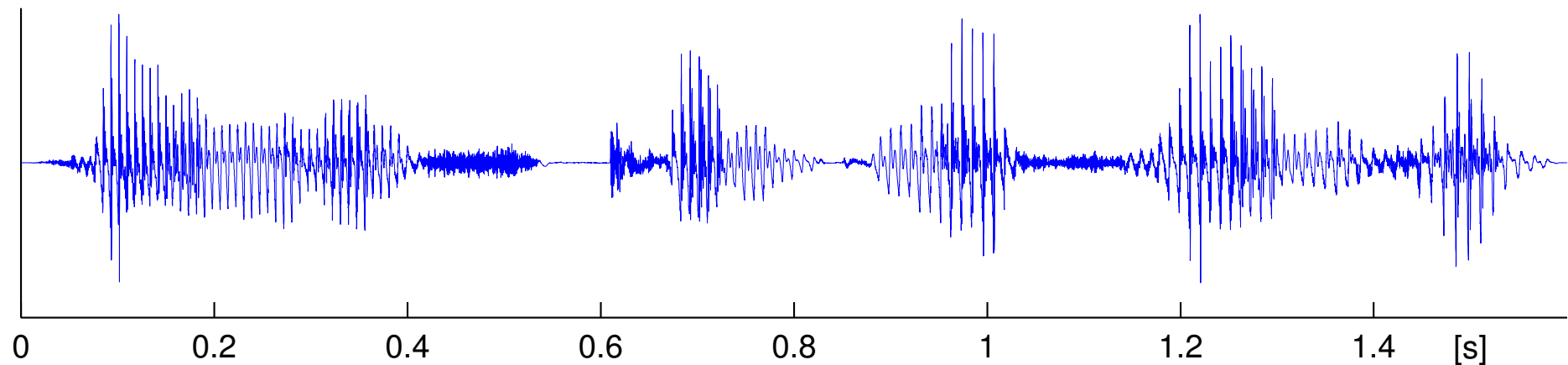
Maschine als Zuhörer: **Spracherkennung** Lautsprache → Text

→ Kenntnis des Zusammenhangs zwischen geschriebener und gesprochener Sprache ist Voraussetzung!

# Geschriebene vs. gesprochene Sprache

Text: "Heinrich kommt nach Hause."

Sprachsignal: 



# Geschriebene vs. gesprochene Sprache

- Text:
- Sequenz diskreter Elemente → **Buchstaben** (Symbole)
  - fester Satz von Symbolen (Alphabet)
  - Elemente sind voneinander unabhängig
- Signal:
- ★ Sequenz von Elementen → **Laute**
  - ★ Laute haben keine scharfen Grenzen
  - ★ stetige Veränderung der Charakteristik
  - ★ benachbarte Laute beeinflussen sich gegenseitig
  - ★ Laute sind von der Stellung abhängig
  - ★ Vokaltrakt prägt die Laute stark
  - ★ Sprechgewohnheiten und Dialekteinfluss



# Beschreibung von Sprache

Sprachsynthese und Spracherkennung setzen voraus, dass die Sprache beschrieben werden kann.

Insbesondere ist auch der **Zusammenhang zwischen geschriebener und gesprochener Sprache wichtig!**

Frage: Wie sieht eine solche Beschreibung aus?

→ **Ist ein wesentliches Thema dieser Vorlesung!**

# Beschreibung aus Sicht der Linguistik

- Graphemische Ebene: Graphem-Inventar >>>  
(geschriebene Sprache)
- Phonemische Ebene: Phonem- und Lautinventar >>>  
(Lautsprache)
- Morphologische Ebene: - Morphem-Inventar >>>  
- Aufbau von Wörtern
- Syntaktische Ebene: Aufbau von Sätzen
- Semantische Ebene: Bedeutung von Wörtern und Sätzen
- Pragmatische Ebene: Zweck und Wirkung der Sprache

# Die akustisch-phonetische Ebene

## Phonetik

Eigenschaften der Laute, Gemeinsamkeiten und Unterschiede  
(im Gegensatz zu Symbolen und Symbolfolgen)

Eigenschaften beziehen sich auf

- das Sprachsignal: periodisch, rauschartig, (quasi-)stationär, ...
- den Sprechapparat: stimmhaft, stimmlos, gerundet, nasal, ...
- die Wahrnehmung: hell, dunkel, weich, hart, ...
- ling. Konventionen: vokalisches, konsonantisches, ...

# Die Prosodie

Laute: kleinste Einheiten der Lautsprache (segmentale Komponente)

Prosodie: lautübergreifende oder suprasegmentale Komponente

Sprechmelodie, Sprechrhythmus, Betonungen ...

→ Eigenheit der gesprochenen Sprache

Im Sprachsignal messbare physikalische prosodische Grössen:

- zeitlicher Verlauf der Grundfrequenz (Tonhöhe)
- zeitlicher Verlauf der Signalintensität
- Dauer der Laute und Pausen

# Die menschliche Sprachproduktion

An der Sprachproduktion beteiligte Organe:

Lunge, Luftröhre, Stimmlippen, Zunge, ...

—> Sprechapparat

>>>

Interessant ist für uns nicht der physiologische Aufbau des Sprechapparates, sondern seine **Funktion**

# Die Funktion des Sprechapparates

## 1. Schallproduktion

- ausgestossene Luft bringt die Stimmlippen in Schwingung
- Luftturbulenzen an Engstellen

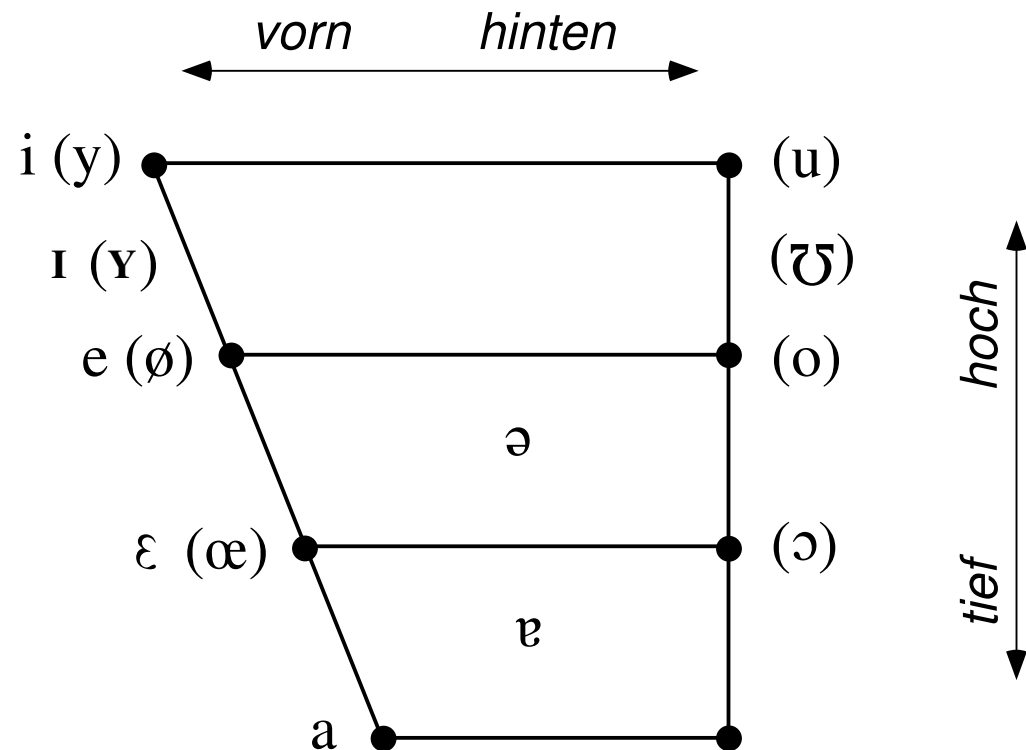
## 2. Klangformung

- neutrales akustisches Signal von den Stimmlippen wird durch Vokaltrakt (Rachen, Mund- und Nasenraum) gefiltert je nach Stellung der Artikulatoren
- Ort der Engstelle bestimmt spektrale Form des Rauschens

# Artikulation der Sprachlaute: Vokale

Das Vokalviereck zeigt den Zusammenhang zwischen der Stellung des Zungenrückens und dem produzierten Laut

(Klammern bedeuten gerundete Laute)



# Konsonanten

Unterteilung nach  
 – Artikulationsart  
 – Artikulationsort

Art \ Ort	Ort							
	bilabial	labiodental	alveolar	palatoalveolar	palatal	velar	dorsouvular	glottal
plosiv	p b		t d			k g		
nasal	m		n			ŋ		
frikativ		f v	s z	ʃ	ç j	x		
lateral			l					
vibriert			r				R	
aspiriert								h



# Das menschliche Gehör

Anatomie des menschlichen Gehörs gibt wenig Aufschluss darüber, wie wir Sprachlaute wahrnehmen.

Wichtiger sind hier **psychoakustische Phänomene** (subj. Wahrnehmung)

- Wahrnehmung:
- Tonhöhe und Periodizität >>>
  - Schallintensität >>>
  - Phase
  - zusammengesetzte Klänge (Verdeckungseffekt) >>>

# Wahrnehmung der Sprachlaute

In der Phonetik werden die Vokale nicht nur nach artikulatorischen Kriterien eingeteilt, sondern auch nach akustischen:

## Formantfrequenzen

- Abhängig von der Stellung der Artikulatoren
- Vokaltrakt hat stets mehrere Resonanzfrequenzen (im Spektrum gut sichtbare relative Maxima)
- Werden mit steigender Frequenz nummeriert:  $F_1$ ,  $F_2$ ,  $F_3$ , ...
- $F_1$  und  $F_2$  sind am wichtigsten für die Lautwahrnehmung  
→ Formantkarte

>>>

# Sprachverarbeitung

Feststellung: Mensch produziert und versteht Lautsprache ohne nennenswerte Anstrengung

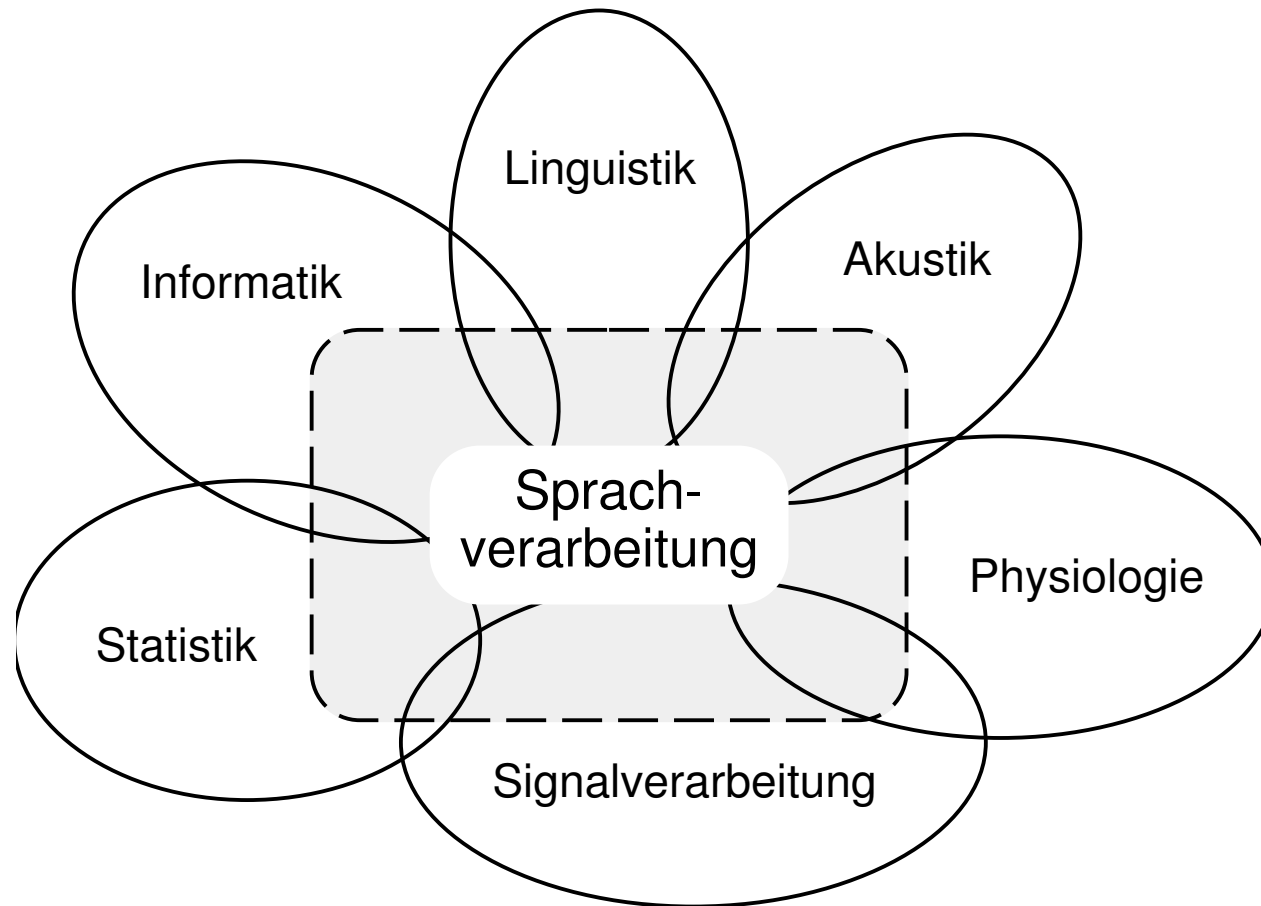
Vermutung: einfache Aufgabe, auch für einen Computer

Realität: Das mit Computer bisher Erreichte ist im Vergleich zu den Fähigkeiten des Menschen bescheiden!

Hauptgründe:

- Mensch setzt enorm viel Wissen und Erfahrung über die Sprache, die Situation, das Thema etc. ein!  
→ Komplexität ist riesig!
- Aufgabe ist interdisziplinär!

# Die Sprachverarbeitung ist interdisziplinär



# Zusammenfassung

- Sprache ist schwierig zu beschreiben (viele Betrachtungsweisen)
- Geschriebene und gesprochene Sprache sind sehr unterschiedlich
- Zusammenhang zwischen Text und Lautsprache ist komplex

Trotzdem fällt dem Menschen das Sprechen und Verstehen leicht!

Was kann die Sprachverarbeitung?

—→ Das vernehmen Sie in dieser Vorlesung!

## Thema der nächsten Lektion:

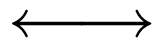
- Übersicht über die Sprachverarbeitung
- Darstellung und Eigenschaften von Sprachsignalen

Zur Übersicht der Vorlesung *Sprachverarbeitung I* >>>





# Graphem



# Graph oder Buchstabe

kleinste Einheit der Schriftsprache

Deutsch:  $\langle a \rangle$ ,  $\langle b \rangle$ ,  $\langle c \rangle$ , ...,  $\langle z \rangle$ ,  
 $\langle ä \rangle$ ,  $\langle ö \rangle$ ,  $\langle ü \rangle$

konkrete Realisierung eines Graphems

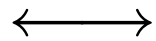
Beispiele: *a*, *b*, *C*, ...

Merke: *a*, *A*, *a* etc. sind Allographen  
des Graphems  $\langle a \rangle$

<<<



# Phoneme



# Phone oder Laute

kleinste **bedeutungsunterscheidende**  
Einheiten der Lautsprache

kleinste Einheiten der Lautsprache

Notation: Lautschrift (IPA)

>>>

/h/ und /m/ sind zwei Phoneme,  
weil sie mindestens ein Wortpaar  
unterscheiden:

Hut /hʊt/ ↔ Mut /mʊt/

[ç] und [x] sind nicht zwei Phoneme,  
weil es kein Wortpaar gibt, welches sich nur  
durch diese Laute unterscheidet

→ *stellungsbedingte Allophone von /x/*  
z.B. sich [sɪç] — noch [nɔx]

auch /a/ und /aː/ sind zwei Phoneme,  
weil sie ein Wortpaar unterscheiden:

Bann /ban/ ↔ Bahn /baːn/

auch [r] und [ʀ] sind nicht zwei Phoneme,  
sie können einander frei ersetzen

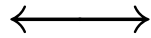
→ *freie Allophone von /r/*

Merke: Zahl der Laute > Zahl der Phoneme > Anzahl der Grapheme!

<<<



# Morphem



# Morph

kleinste **bedeutungstragende**  
Einheiten der Sprache

konkrete Realisierung eines  
Morphems

Hauptsächlich Stämme oder  
Stammmorpheme

Beispiele: “haus”, “baum”, “viel”, ...

Merke: “haus” und “häus” sind Allomorphe  
des Morphems ⟨haus⟩

<<<



# Die phonetische Schrift

Es gibt mehr Laute als Buchstaben

→ Spezieller Symbolsatz für die Notation der Lautsprache nötig!

IPA-Symbole (International Phonetic Association)

Beispiel: “Sprachverarbeitung” [ʃpraːxferarbaɪtʊŋ]

Problem: IPA-Symbole nicht im Computer-Zeichensatz

# ASCII-Repräsentation der IPA-Symbole

IPA	ETHPA	Beispiel	
a	a	hat	[hat]
aː	aː	Bahn	[baːn]
ai	a_i	weit	[vait]
ç	C	sich	[ziç]
eː	eː	Beet	[beːt]
ɛ	E	hätte	['hɛtə]
ɛː	Eː	wähle	['vɛːlə]
ŋ	N	lang	[laŋ]
ʃ	S	schal	[ʃaːl]

(die vollständige Liste ist im Anhang A.1.1 des Buches zu finden)

<<<



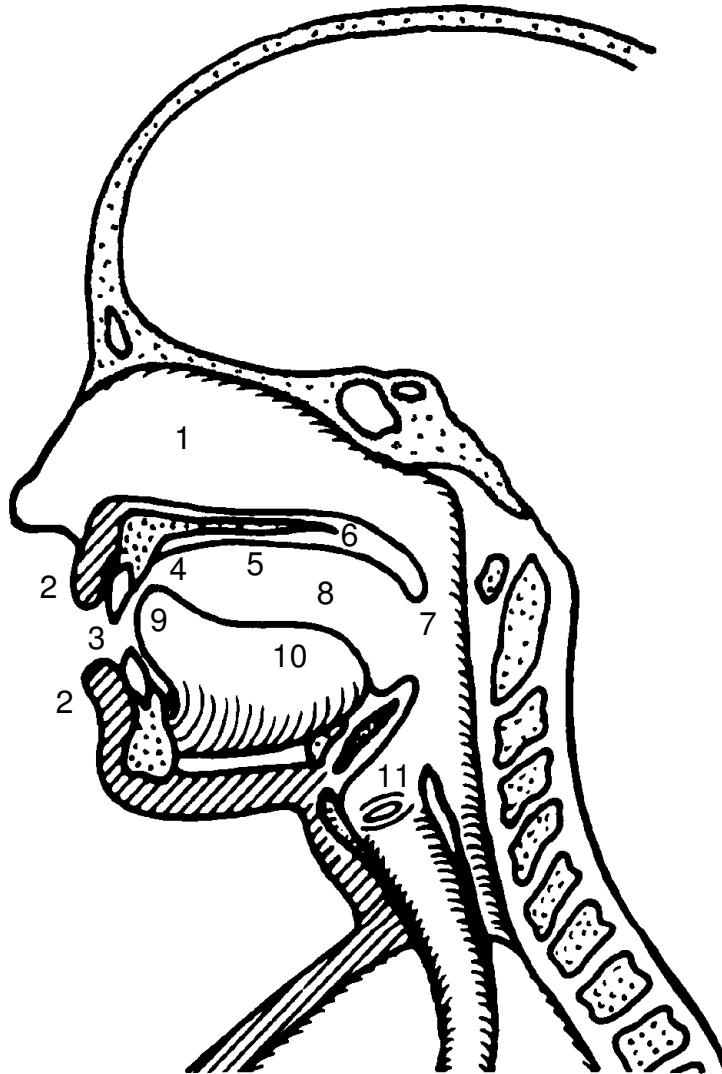


# Was ist die Funktion der Prosodie?

Linguistische Funktion:	Kennzeichnen	(z.B. Satztyp)
	Gewichten	(betonen der wichtigen Teile)
	Gliedern	(gruppieren zusammengehöriger Wörter)
Ausserlinguist. Funktion:	Stimme gehetzt, zaghaft, traurig, wütend, ...	

<<<

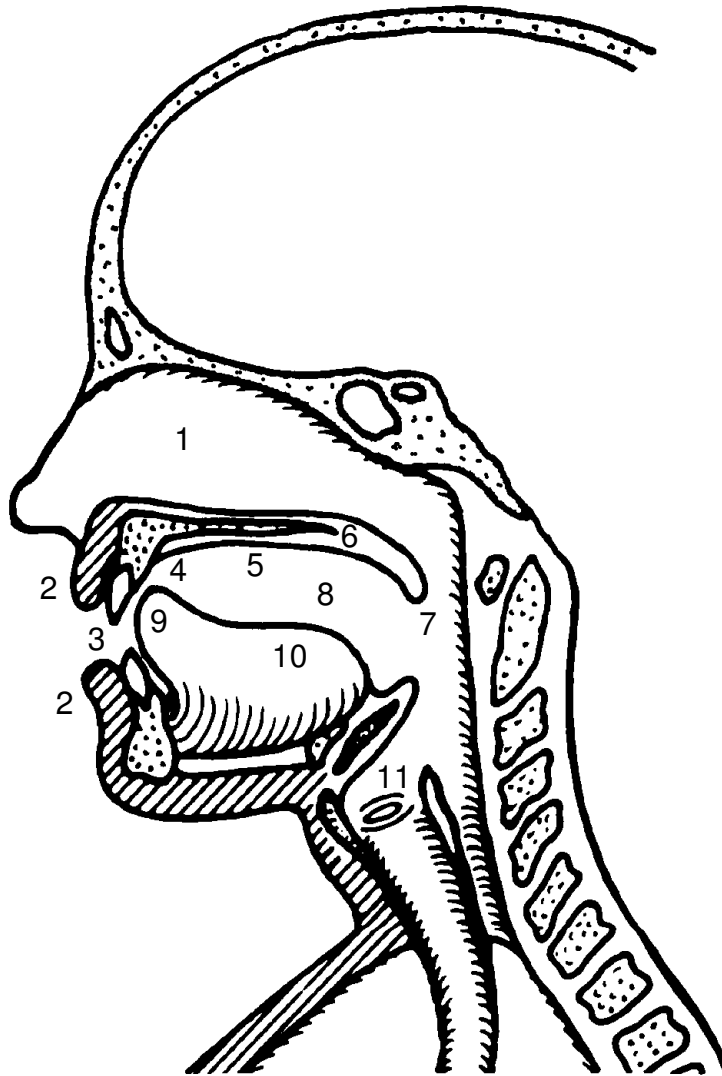




- 1 Nasenraum (Cavum nasi)
- 2 Lippen (Labia)
- 3 Zähne (Dentes)
- 4 Zahndamm (Alveolen)
- 5 harter Gaumen (Palatum)
- 6 weicher Gaumen (Velum)
- 7 Halszäpfchen (Uvula)
- 8 Mundraum (Cavum oris)
- 9 Zungenspitze (Apex)
- 10 Zungenrücken (Dorsum)
- 11 Stimmlippen im Kehlkopf (Larynx)

<<<





- 1 Nasenraum (Cavum nasi)
- 2 Lippen (Labia)
- 3 Zähne (Dentes)
- 4 Zahndamm (Alveolen)
- 5 harter Gaumen (Palatum)
- 6 weicher Gaumen (Velum)
- 7 Halszäpfchen (Uvula)
- 8 Mundraum (Cavum oris)
- 9 Zungenspitze (Apex)
- 10 Zungenrücken (Dorsum)
- 11 Stimmlippen im Kehlkopf (Larynx)

<<<



# Tonhöhe und Periodizität

Periodizität:	periodisch	→ Ton / Klang	>>>
	aperiodisch	→ Rauschen	>>>
Tonhöhe:	Bereich:	ca. 20 Hz – 18 kHz (altersabhängig)	
	Genauigkeit:	ca. 0.1 % @ 1 kHz (frequenzabhängig)	
	rel. Frequenz:	z.B. Halbton $f_2/f_1 = 2^{1/12}$	>>>
			<<<



# Töne vs. Klänge

Ton:

Sinussignal

$$f = 220 \text{ Hz}$$

2 Sinussignale

$$f_1 = 220 \text{ Hz} \quad (\text{a})$$

$$f_2 = 277 \text{ Hz} \quad (\text{cis'})$$

Klang:

harmonisches Signal

$$F_0 = 220 \text{ Hz}$$

(36 Sinussignale)

2 harmonische Signale

$$F_{0_1} = 220 \text{ Hz}$$

$$F_{0_2} = 277 \text{ Hz}$$

ohne Grundwelle

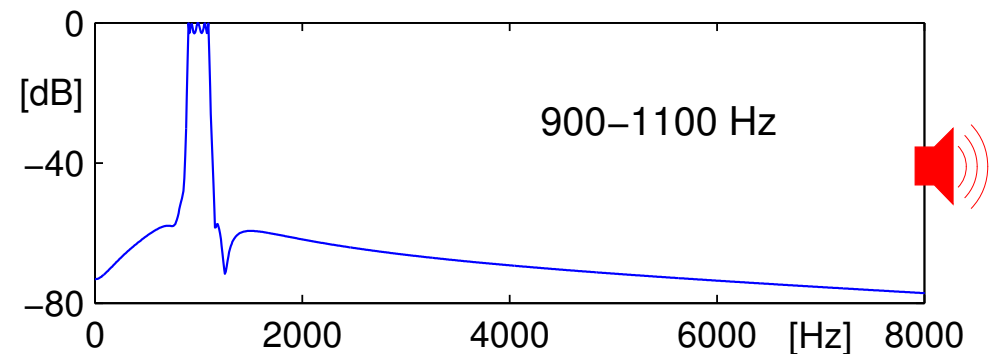
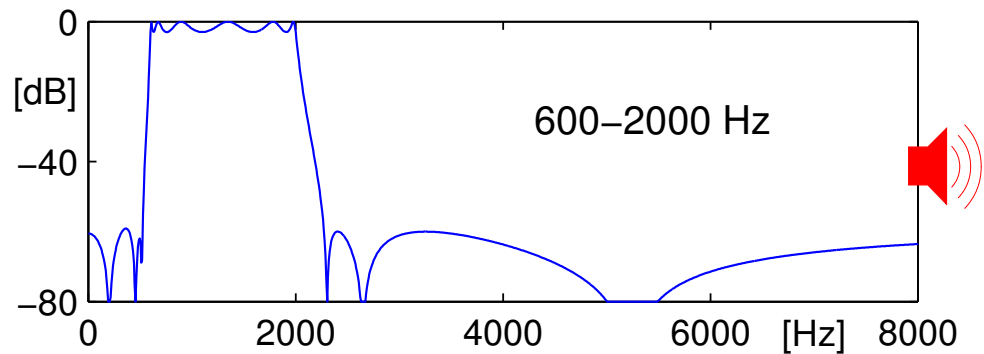
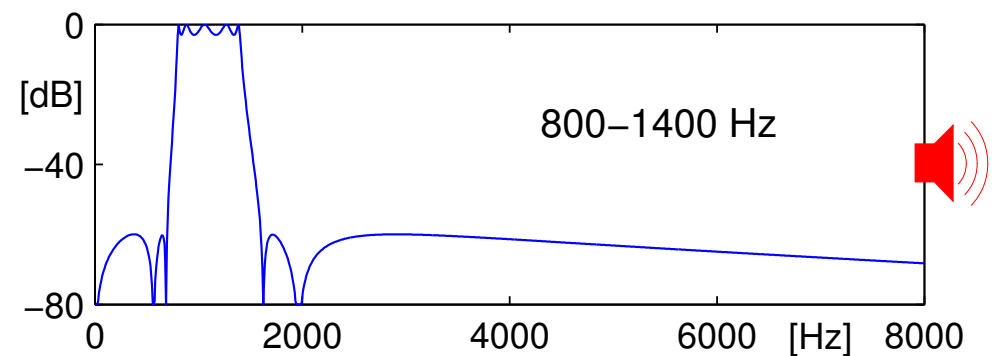
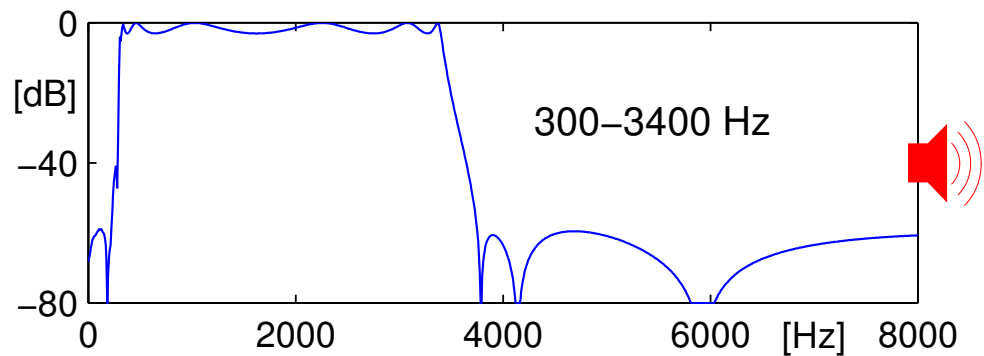
Rauschen:

weisses Rauschen

<<<



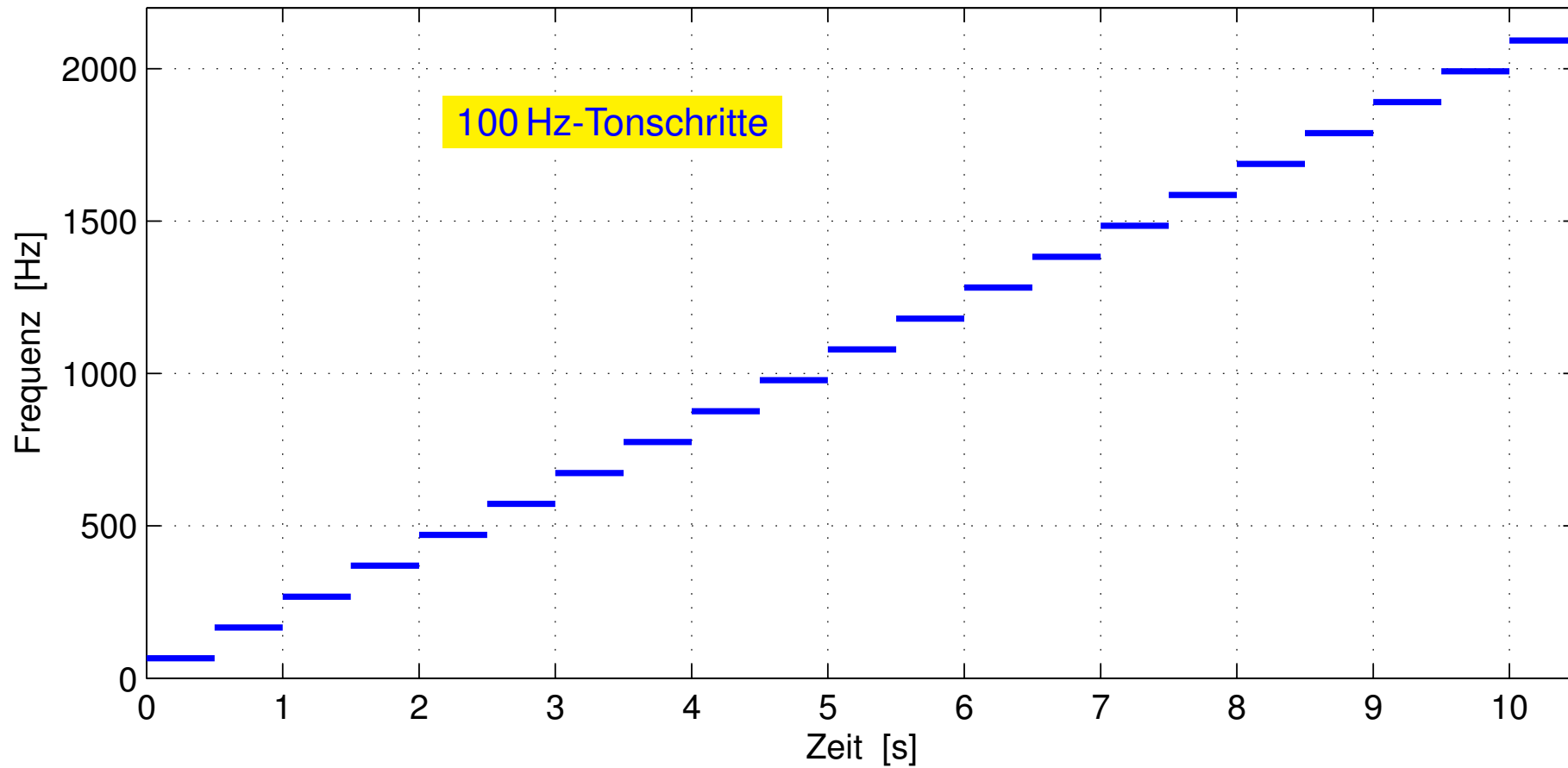
## Mit Bandpässen gefiltertes weisses Rauschen



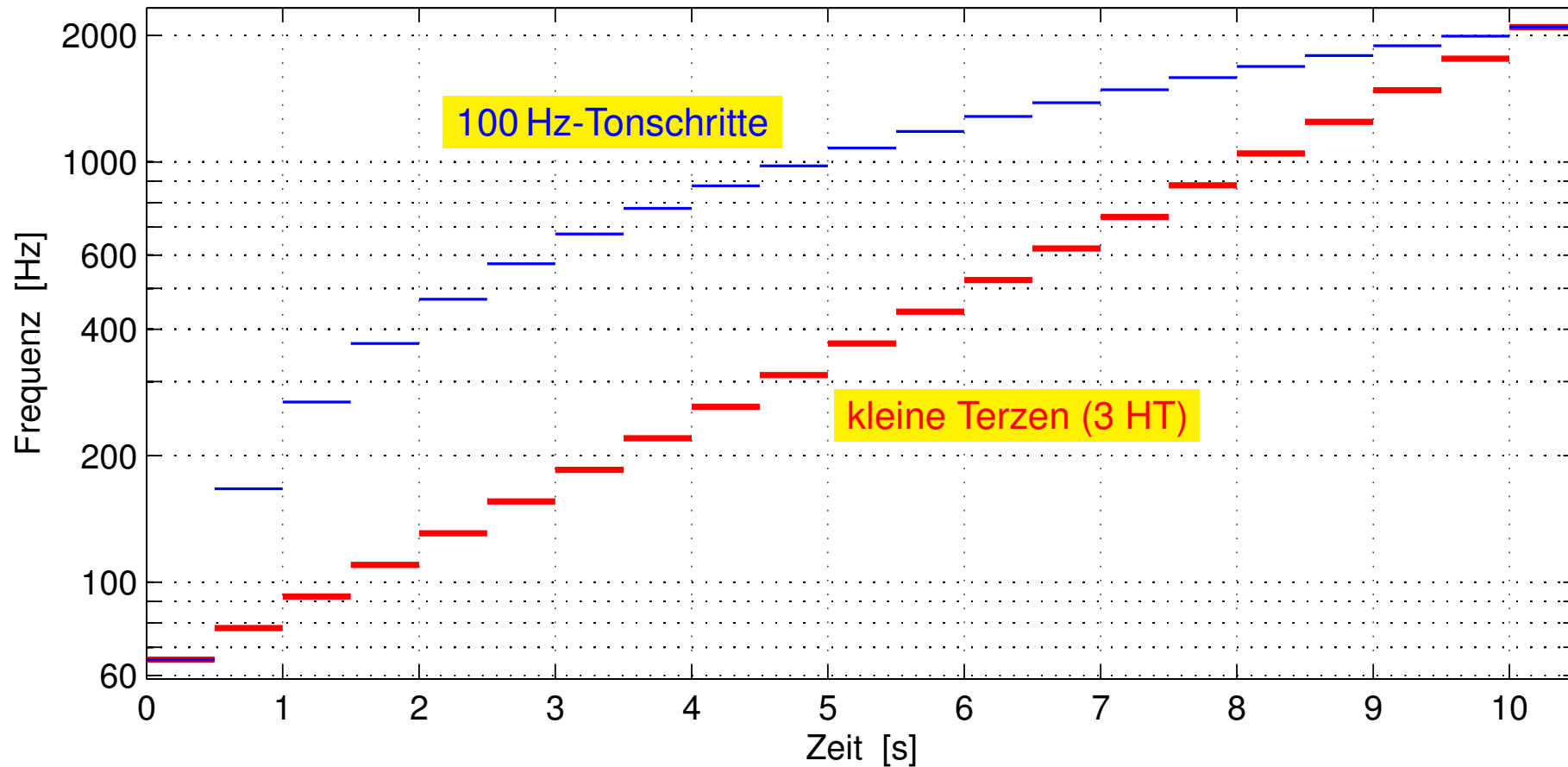
<<<



# Tonschritte auf linearer Frequenzskala

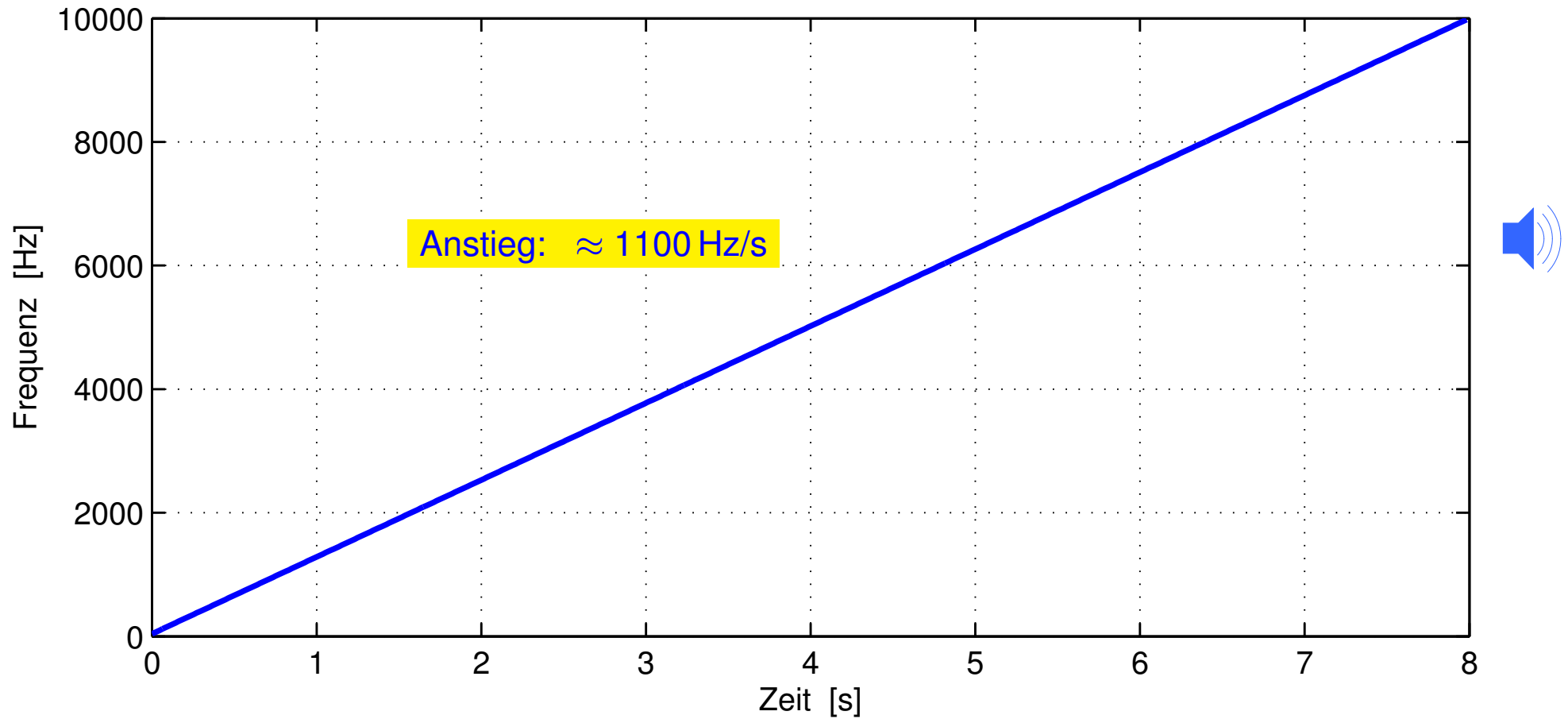


# Tonschritte auf **logarithmischer** Frequenzskala

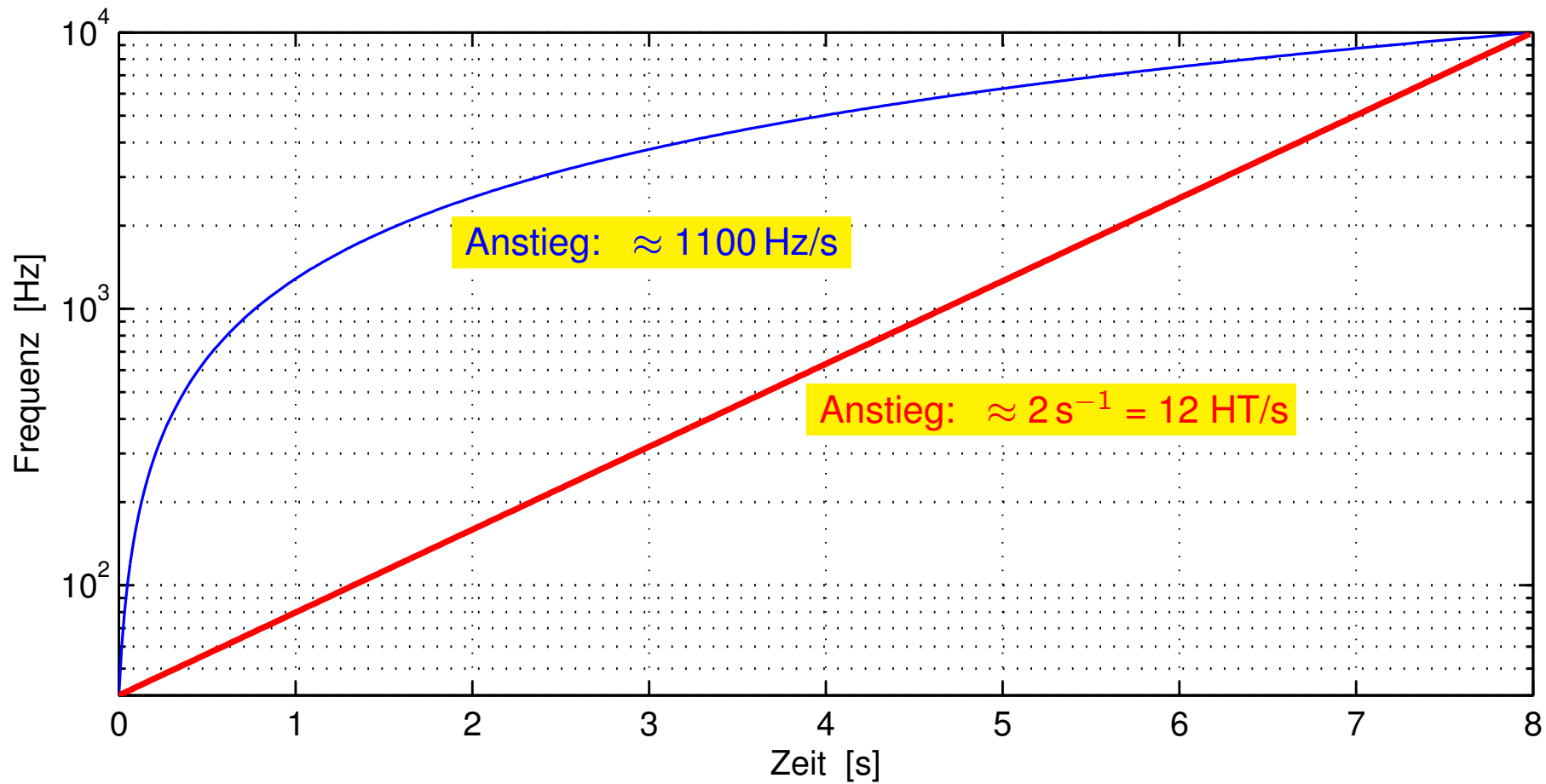


<<<

## Ansteigender Ton auf **linearer** Frequenzskala



# Ansteigender Ton auf **logarithmischer** Frequenzskala



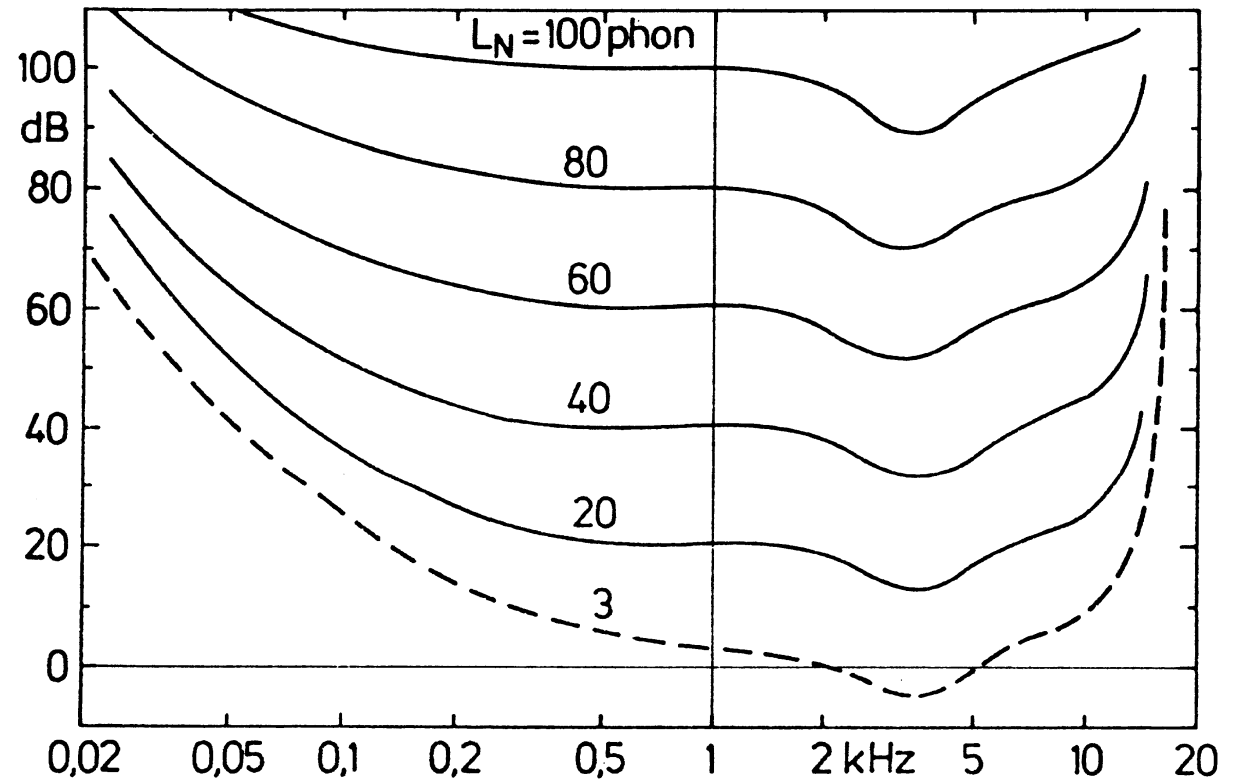
<<<





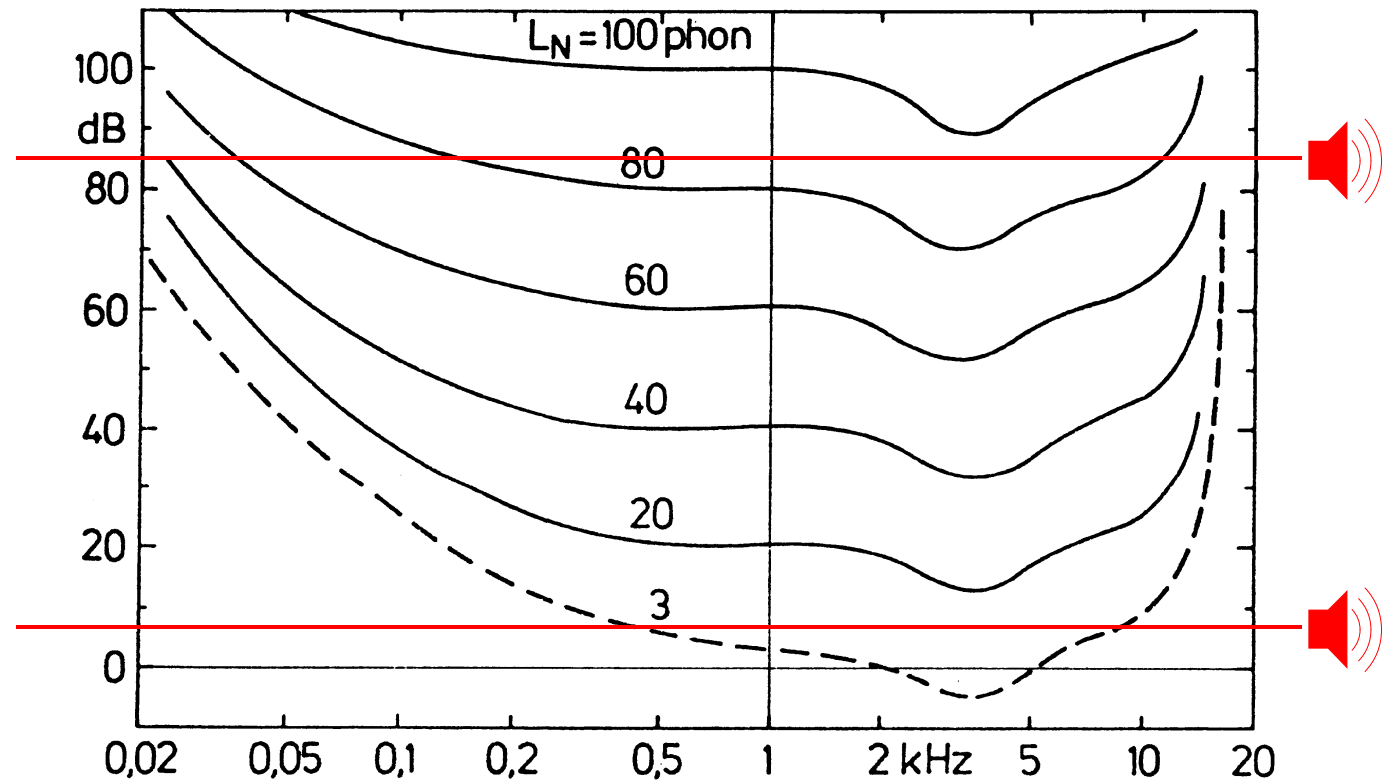
# Wahrnehmung der Schallintensität

Der Hörbereich von der Ruhehörschwelle bis zur Schmerzgrenze umfasst gut 120 dB



# Wahrnehmung der Schallintensität

Der Hörbereich von der Ruheshörschwelle bis zur Schmerzgrenze umfasst gut 120 dB

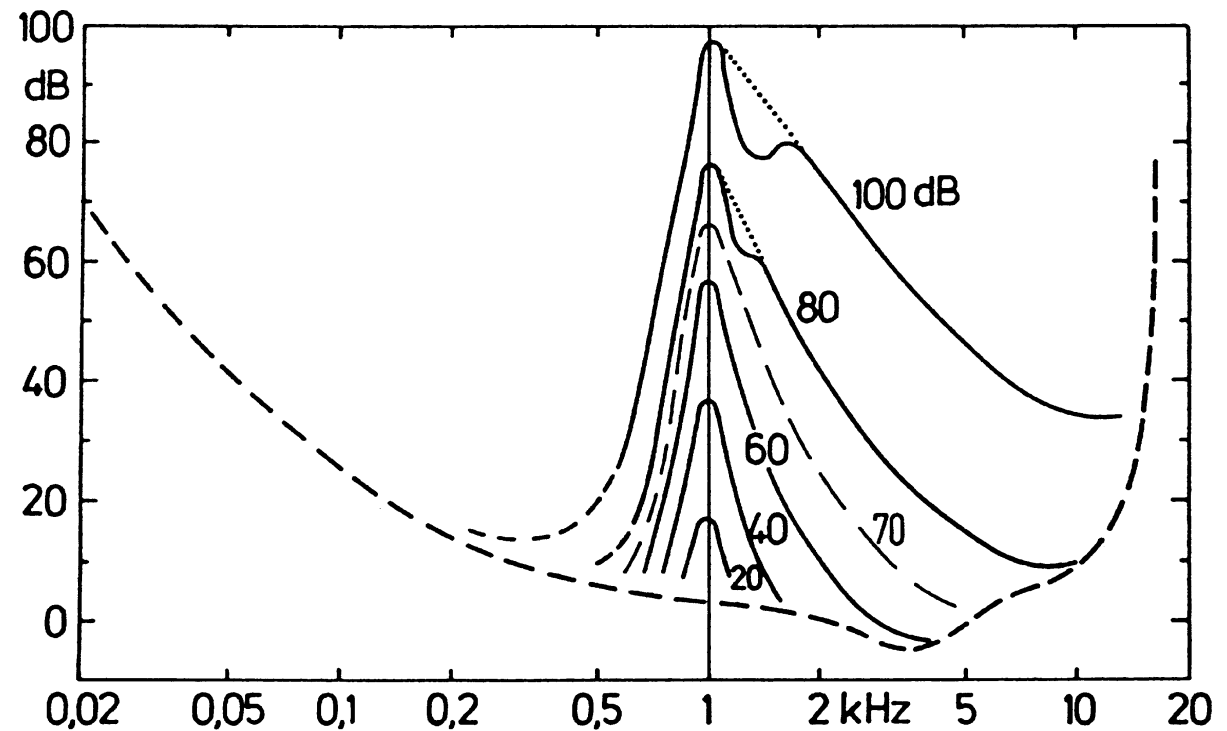


<<<



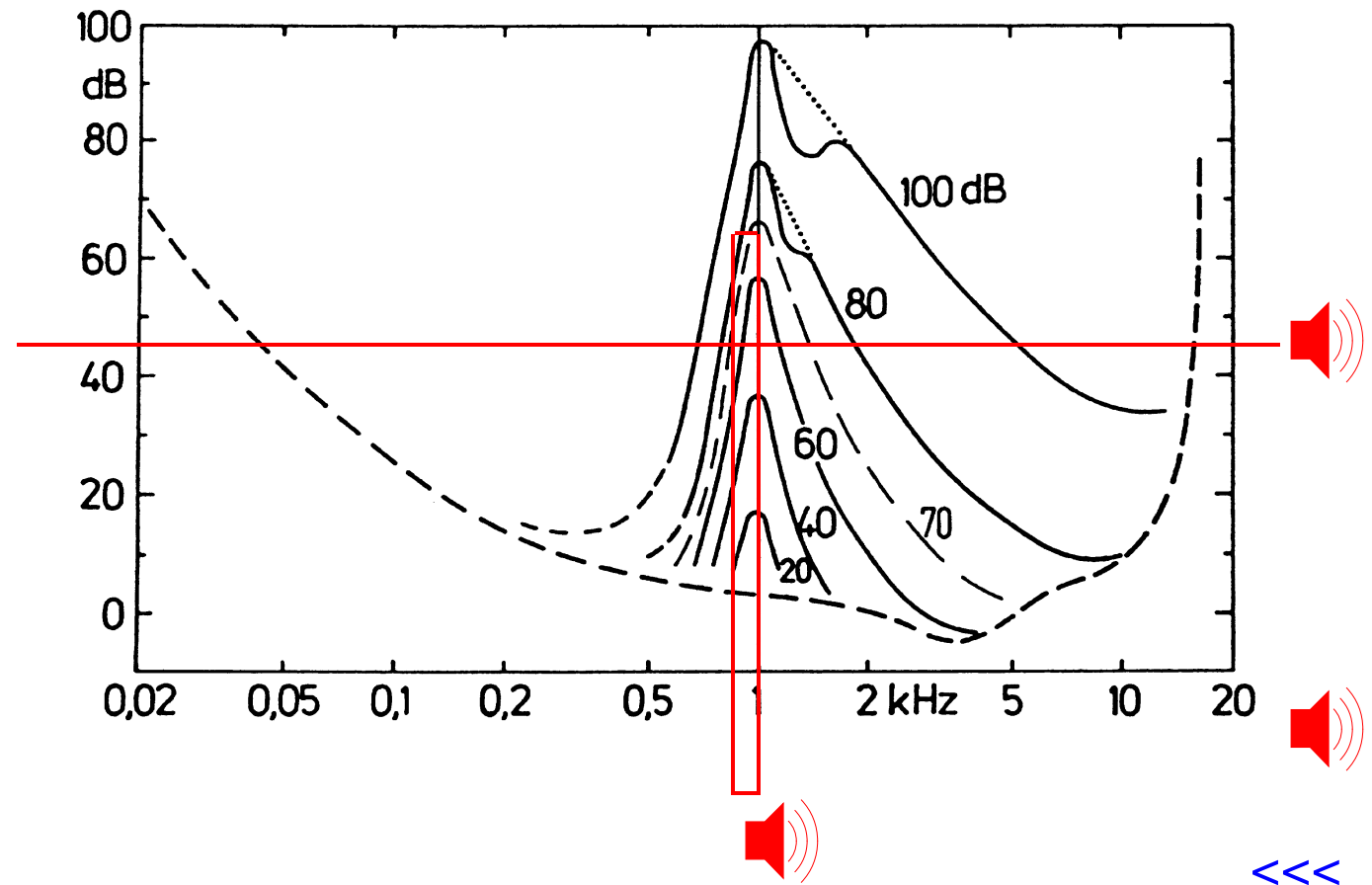
# Der Verdeckungseffekt

Mithörschwelle für  
Schmalbandrauschen  
mit verschiedenen  
Schallpegeln



# Der Verdeckungseffekt

Mithörschwelle für  
Schmalbandrauschen  
mit verschiedenen  
Schallpegeln

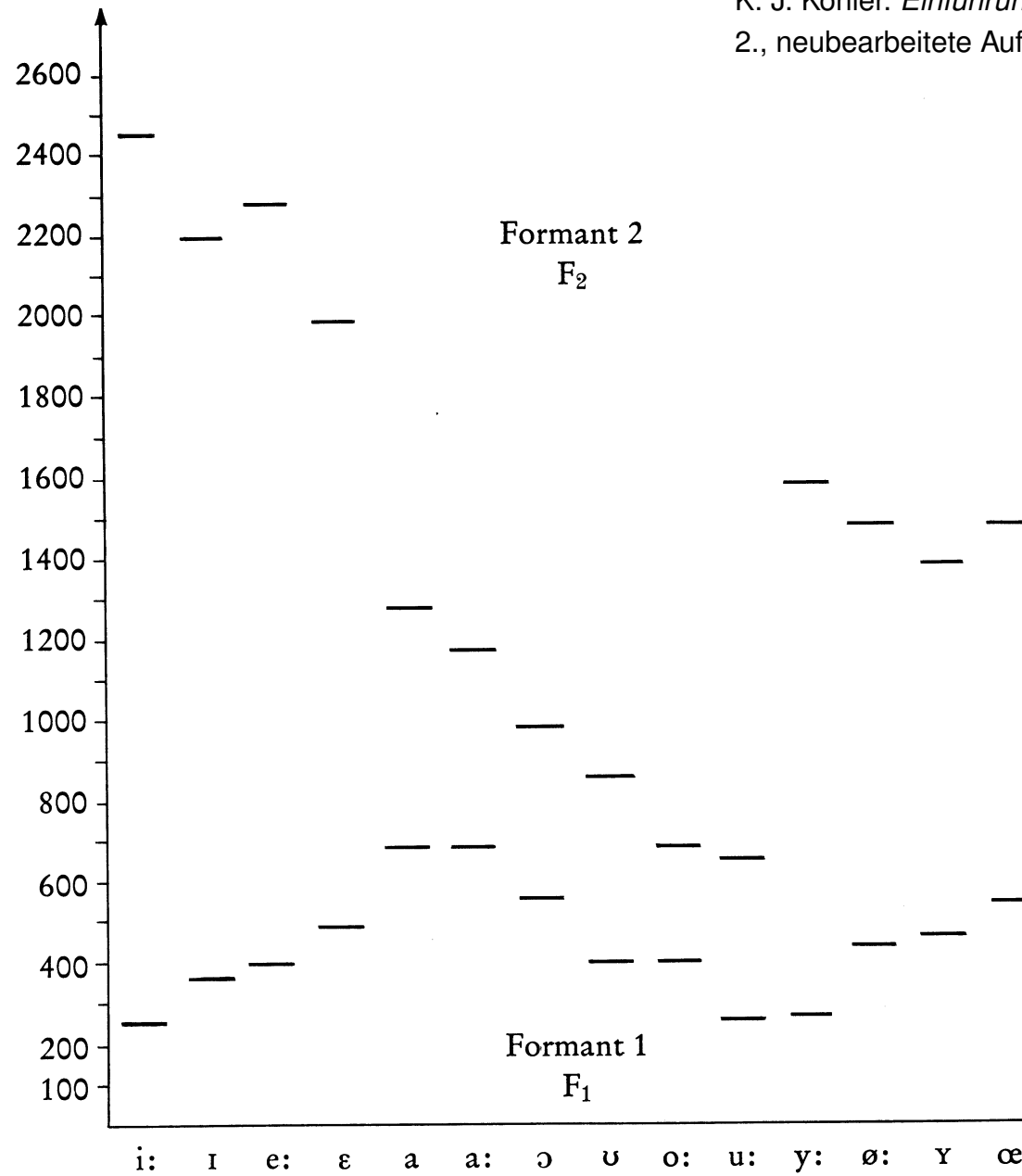




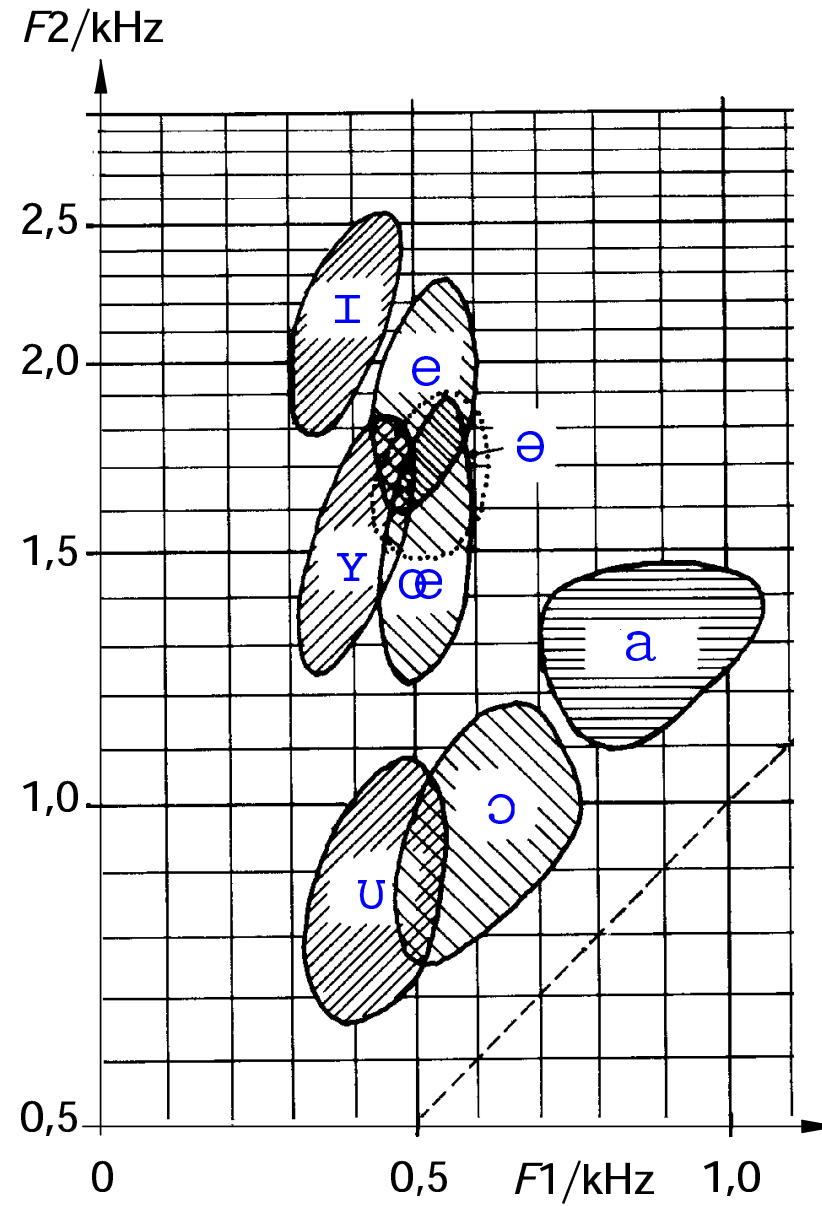
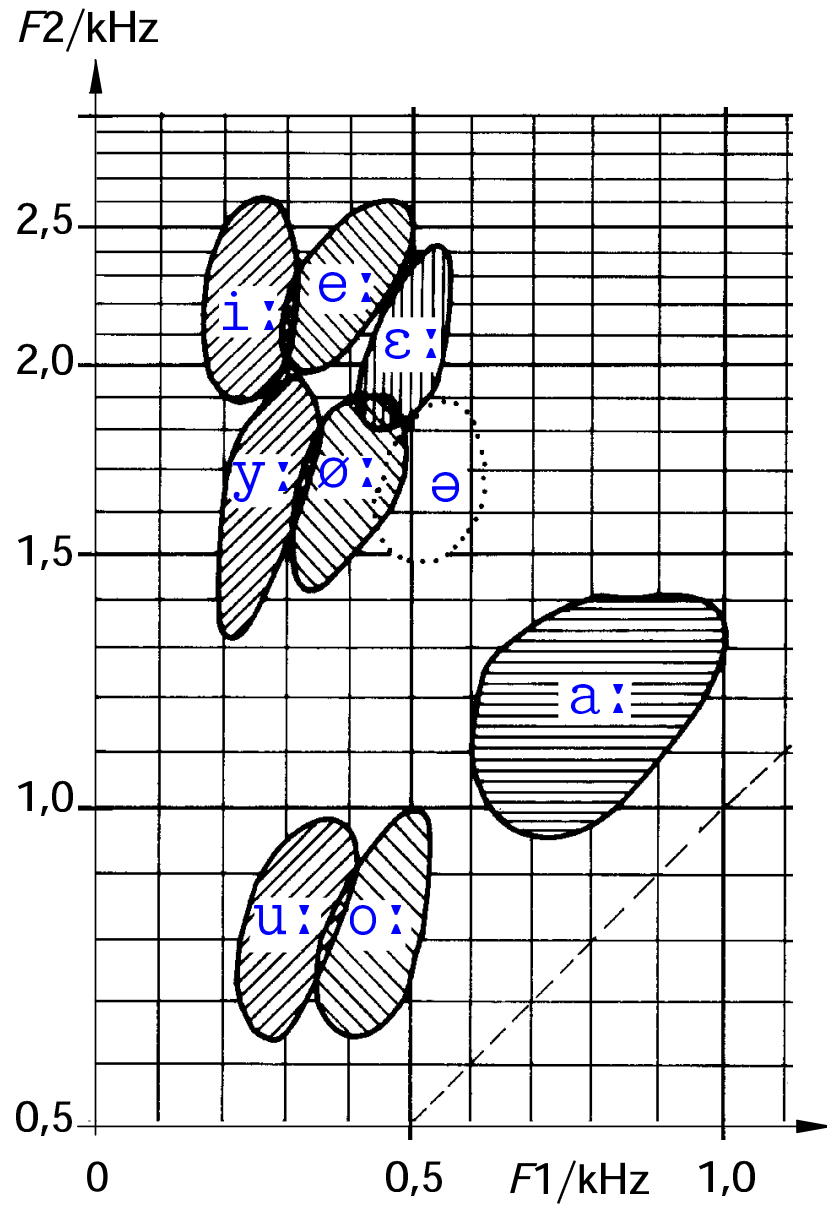
Graphik entnommen aus:

K. J. Kohler. *Einführung in die Phonetik des Deutschen*.

2., neubearbeitete Auflage, Erich Schmidt Verlag, Berlin, 1995.







<<<

