

Kapitel 2: Sprachkommunikation

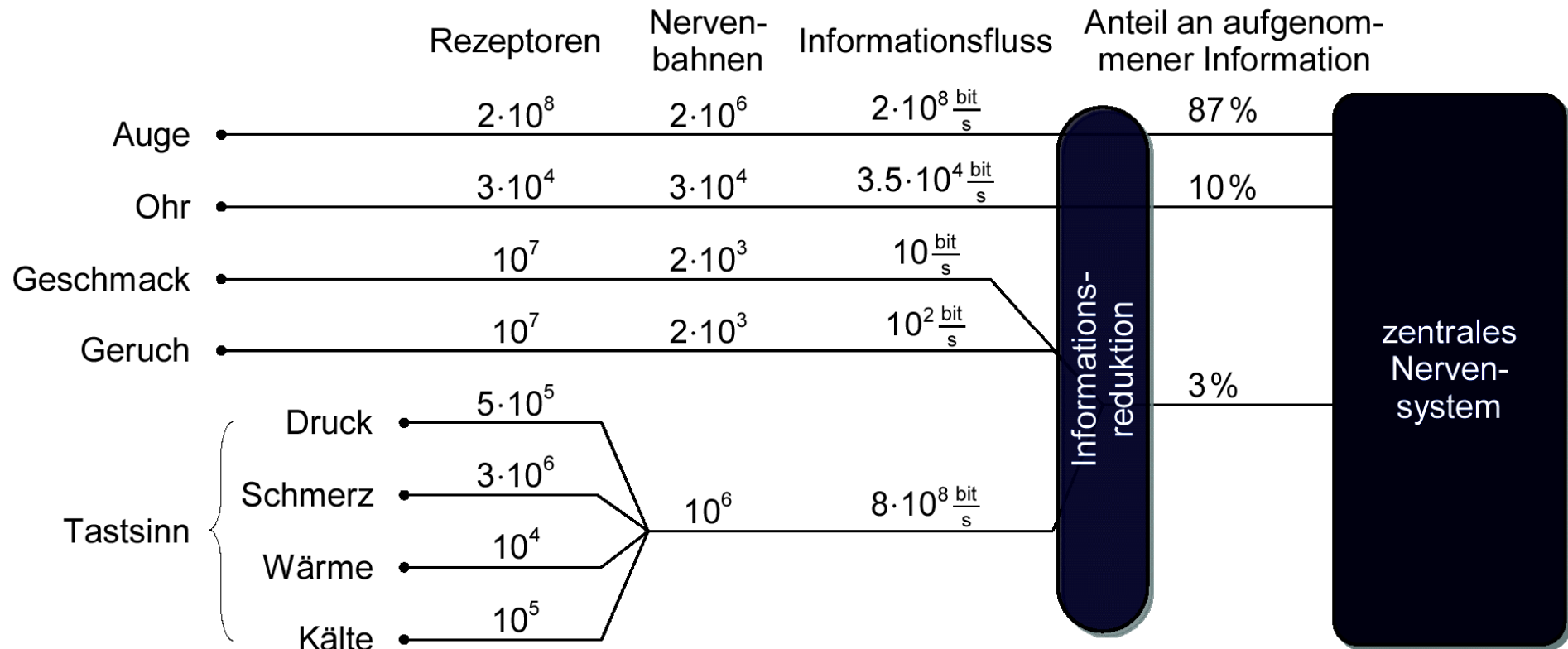
Bedeutung der Sprachkommunikation

- ✓ User-Interface verantwortlich für Kommunikationsfähigkeit des Systems
- ✓ wichtigste Komponenten eines interaktiven Systems:
 - Dialogkomponente (für Navigation im Menü)
 - Benutzerschnittstelle (für Ein-/Ausgabe)
- ✓ hierfür verschiedene Modalitäten möglich (Standard: Bildschirm, Tastatur, Maus)
- ✓ fortgeschrittene Ein-/Ausgabemethoden: Sprache, Handschrift, Gestik, Zeigen
- ✓ Sprachkommunikation: Hat größtes Potential aller Eingabemethoden, da auch beim Menschen die häufigste und natürlichste Kommunikationsform
- ✓ Sprachkommunikation: Ermittlung der geäußerten Wortfolge aus einem vorliegenden Sprachsignal und Verarbeitung dieser Information

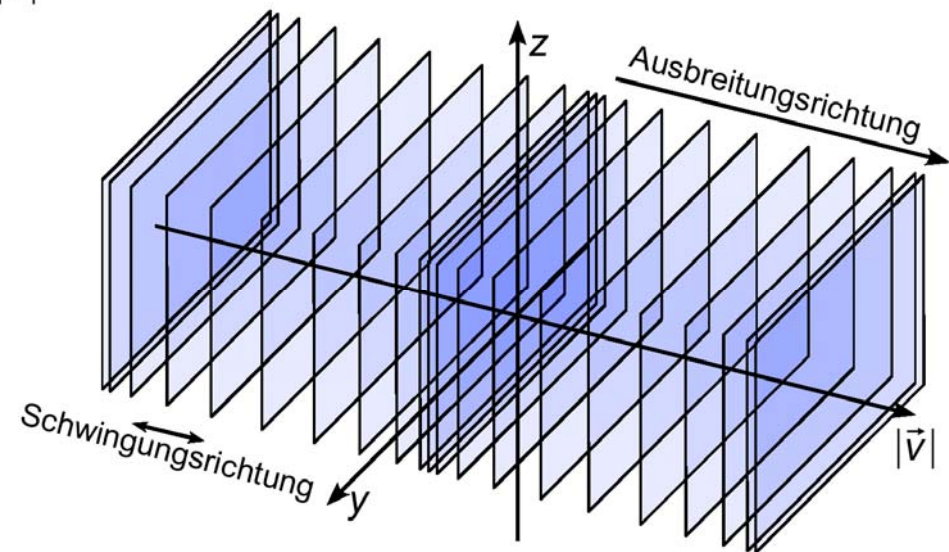
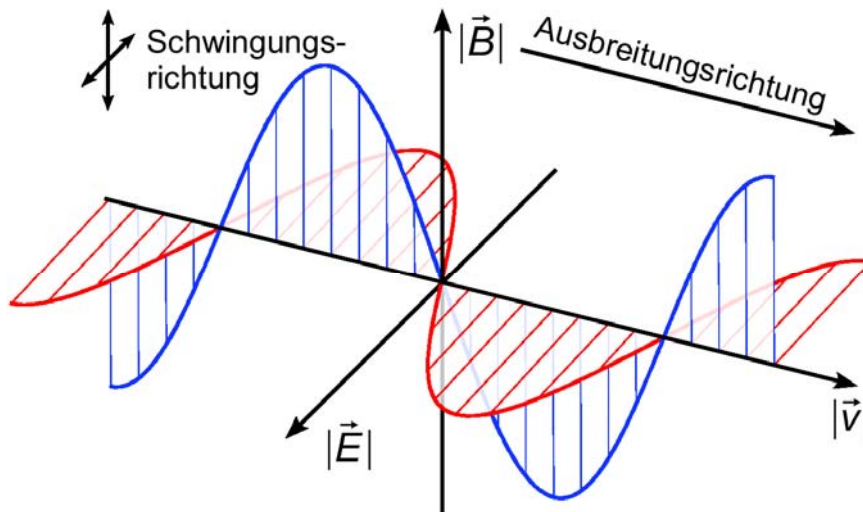
Datenraten gängiger Systeme der MMK

<u>System</u>	<u>Verhalten</u>	<u>Rate (KByte/sec)</u>
Tastatur (ungeübt)	Eingabe	0.01
Tastatur (geübt)	Eingabe	0.025
Handschrift	Eingabe	0.0025
Spracheingabe	Eingabe	0.01-0.02
Maus	Eingabe	0.02
Sprachausgabe	Ausgabe	0.6
Text lesen	Ausgabe	0.03-0.3
Hören (CD)	Ausgabe	40
Sehen (Video)	Ausgabe	20 000

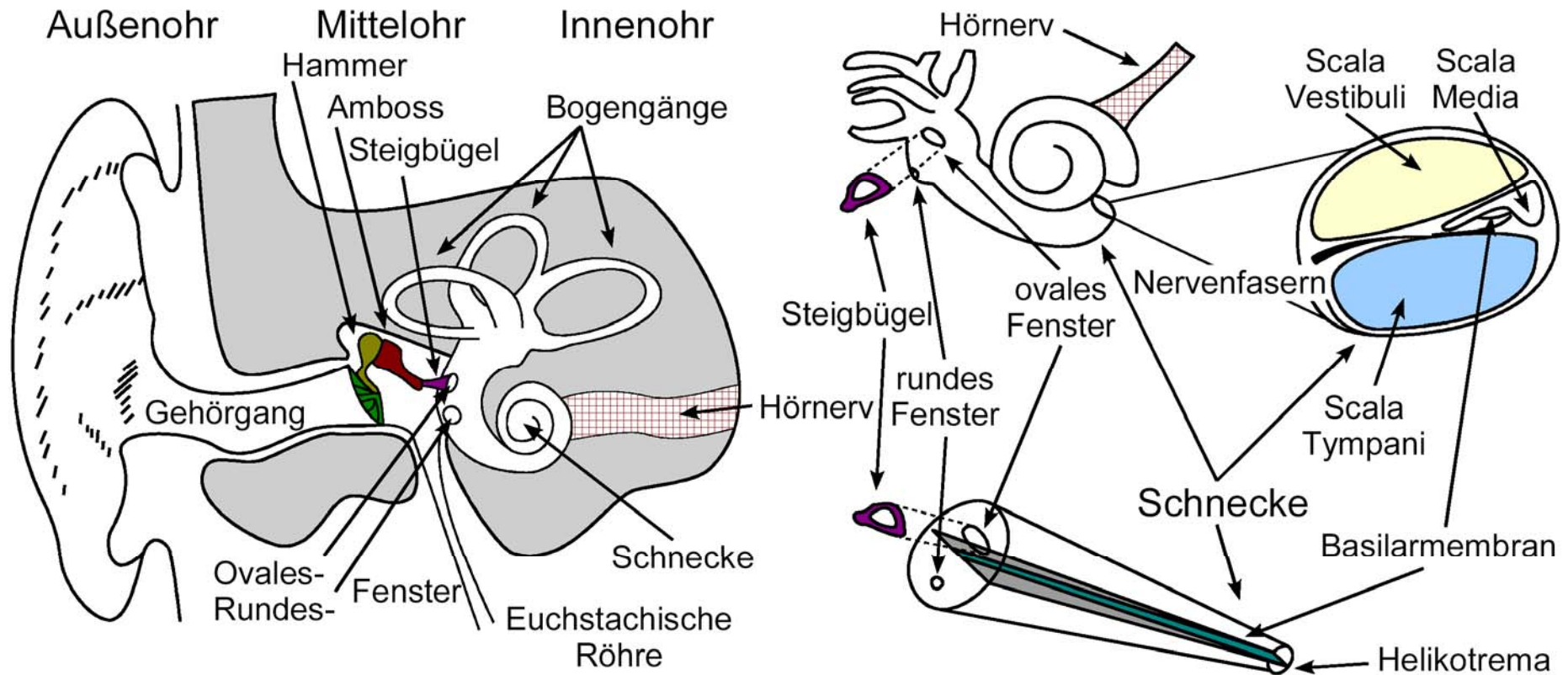
Die Sinne des Menschen und ihre Datenraten



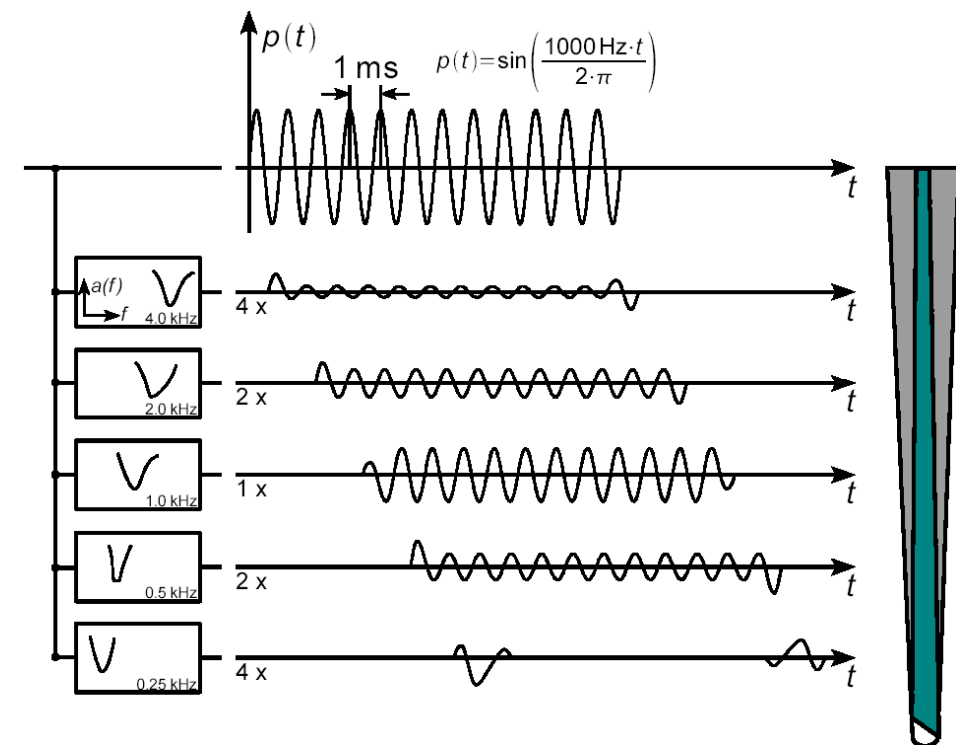
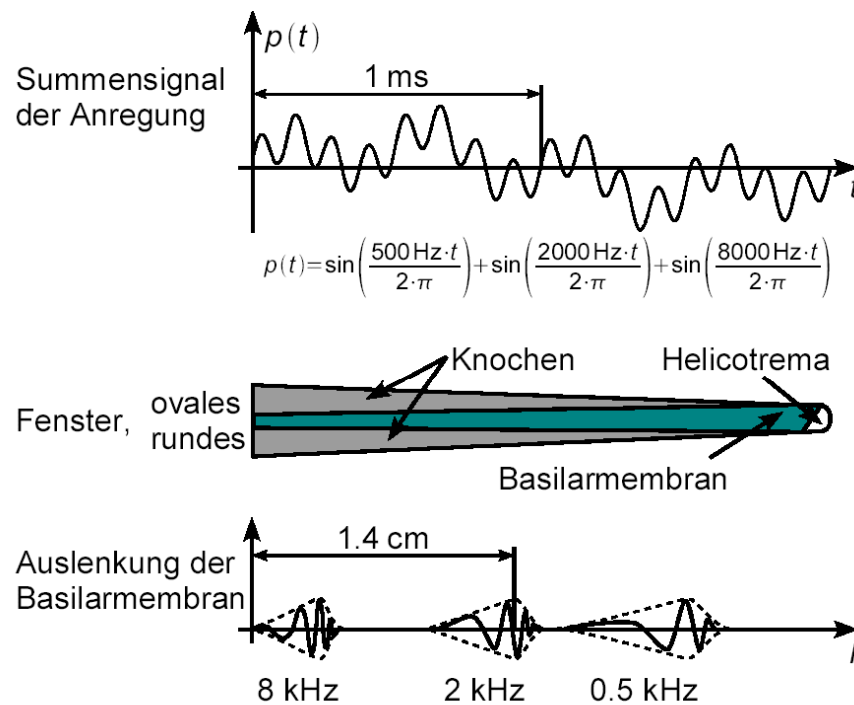
Transversal- und Longitudinalwelle



Schematische Darstellung des Ohrs



Frequenzempfindlichkeit der Basilar membrane



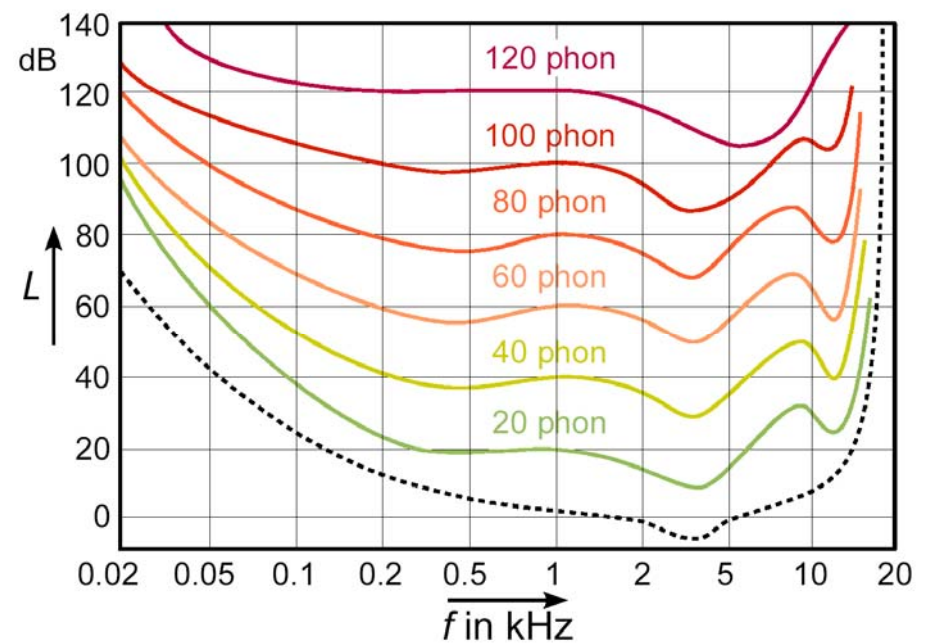
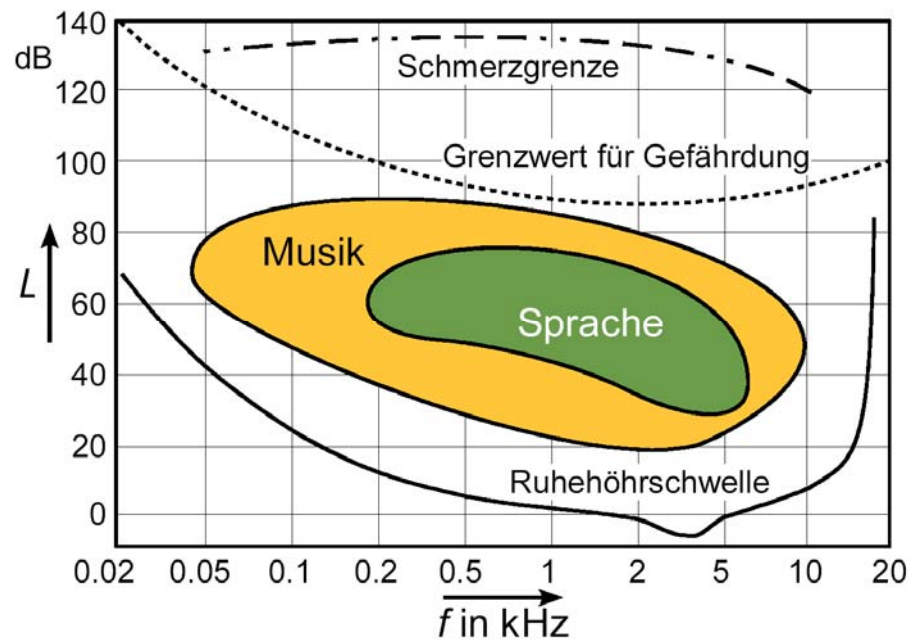
Psychoakustische Meßgrößen

Psychoakustik		Physik	
Bezeichnung	Einheit	Bezeichnung	Einheit
Tonheit Z	Bark	Frequenz f	Hz
Verhältnistonhöhe V	mel	Schalldruck p	$\frac{\text{N}}{\text{m}^2} = \frac{\text{Ws}}{\text{m}^3} = \text{Pa}$
		Schallschnelle v	$\frac{\text{m}}{\text{s}}$
		Schallintensität I	$\frac{\text{W}}{\text{m}^2} = \frac{\text{N}}{\text{s} \cdot \text{m}}$
Lautstärkepegel L_n	Phon	Schalldruckpegel L	dB
Lautheit N	sone	Schallleistung P_{ak}	$\text{W} = \frac{\text{N} \cdot \text{m}}{\text{s}}$
Bezugsschalldruck $p_0 = 2 \cdot 10^{-5} \frac{\text{N}}{\text{m}^2} = 20 \mu\text{Pa}$, Bezugsintensität $I_0 = 1.0 \cdot 10^{-12} \frac{\text{W}}{\text{m}^2}$			

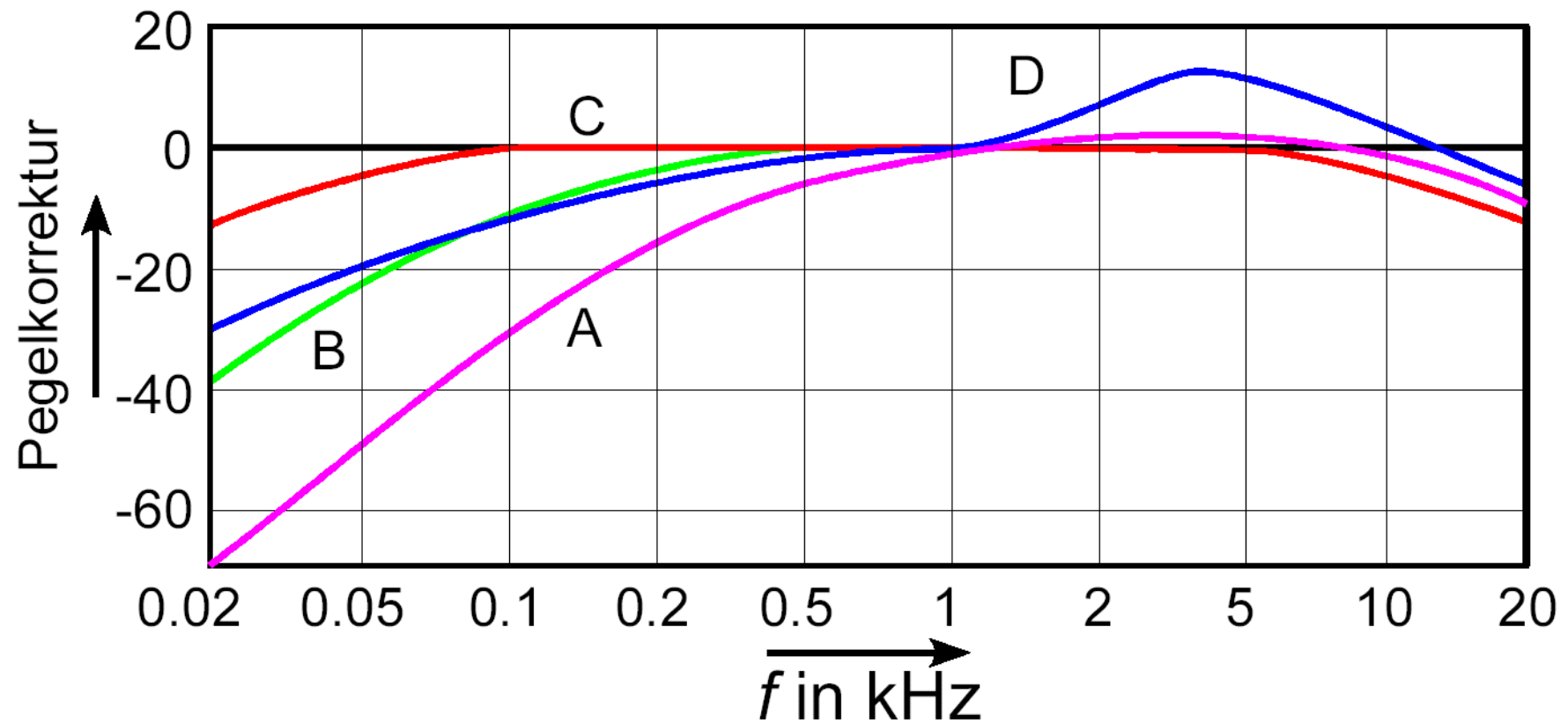
Einige Schallquellen und ihre typischen Pegel

Schall ggf. Entfernungsangabe	Pegel
Düsenjäger, 30 m	$L = 140 \text{ dB} \equiv p = 2.0 \cdot 10^2 \text{ Pa}$
lautes Händeklatschen, 1 m	$L = 130 \text{ dB} \equiv p = 63 \text{ Pa}$
Trillerpfeife, 1 m	$L = 120 \text{ dB} \equiv p = 20 \text{ Pa}$
Discman	$L = 110 \text{ dB} \equiv p = 6.3 \text{ Pa}$
Presslufthammer, 10 m	$L = 100 \text{ dB} \equiv p = 2 \text{ Pa}$
laute Fabrikhalle	$L = 90 \text{ dB} \equiv p = 6.3 \cdot 10^{-1} \text{ Pa}$
starker Straßenverkehr	$L = 80 \text{ dB} \equiv p = 2.0 \cdot 10^{-1} \text{ Pa}$
Staubsauger, 1 m	$L = 70 \text{ dB} \equiv p = 6.3 \cdot 10^{-2} \text{ Pa}$
normale Sprache, 1 m	$L = 60 \text{ dB} \equiv p = 2.0 \cdot 10^{-2} \text{ Pa}$
Kühlschrank, 1 m	$L = 50 \text{ dB} \equiv p = 6.3 \cdot 10^{-3} \text{ Pa}$
normale Wohngeräusche, 1 m	$L = 40 \text{ dB} \equiv p = 2.0 \cdot 10^{-3} \text{ Pa}$
Flüstersprache	$L = 30 \text{ dB} \equiv p = 6.3 \cdot 10^{-4} \text{ Pa}$
mechanischer Wecker, 1 m	$L = 20 \text{ dB} \equiv p = 2.0 \cdot 10^{-4} \text{ Pa}$
Blätterrauschen in der Ferne	$L = 10 \text{ dB} \equiv p = 6.3 \cdot 10^{-5} \text{ Pa}$

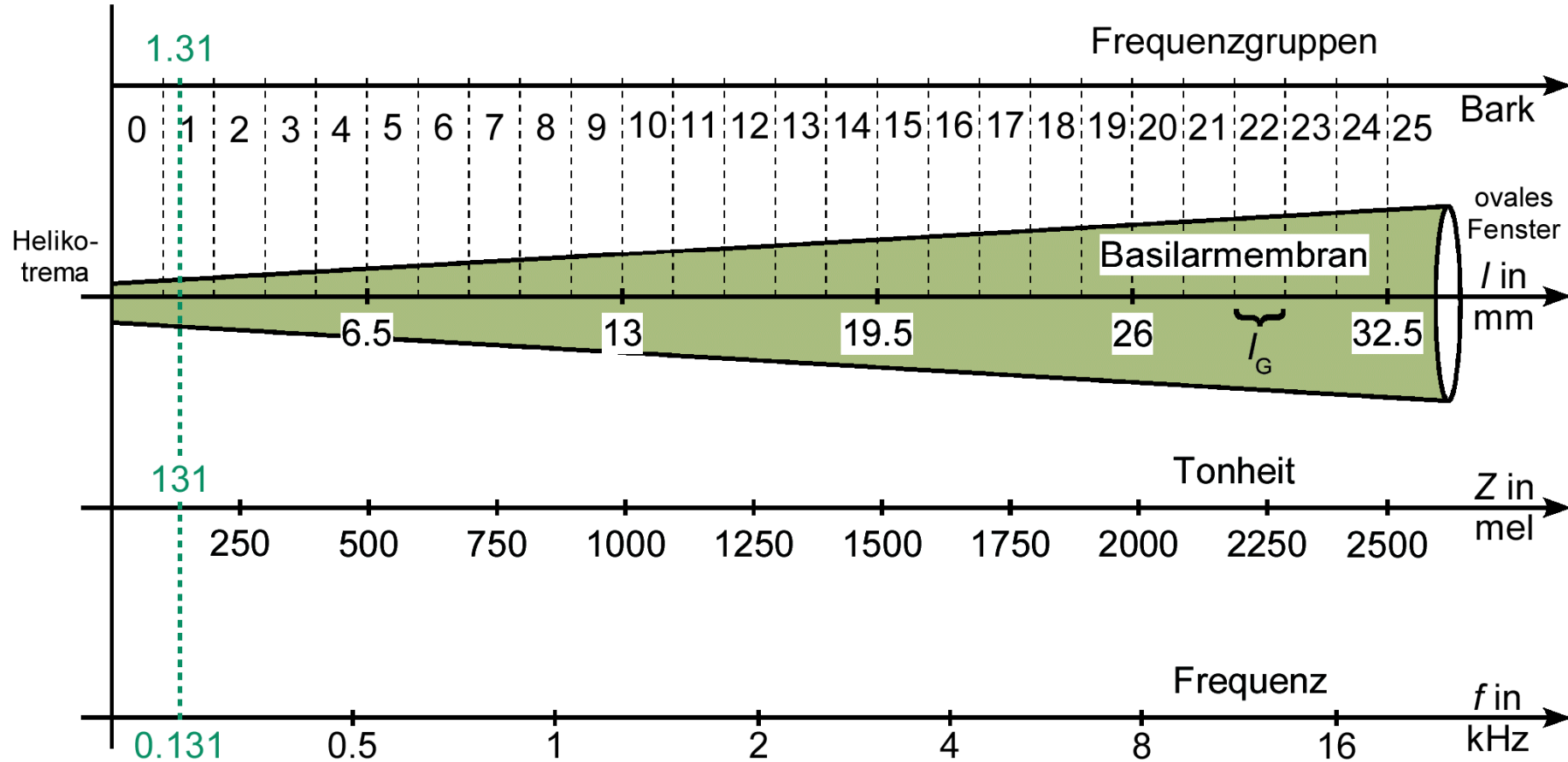
Die Hörfläche nach DIN.45.630



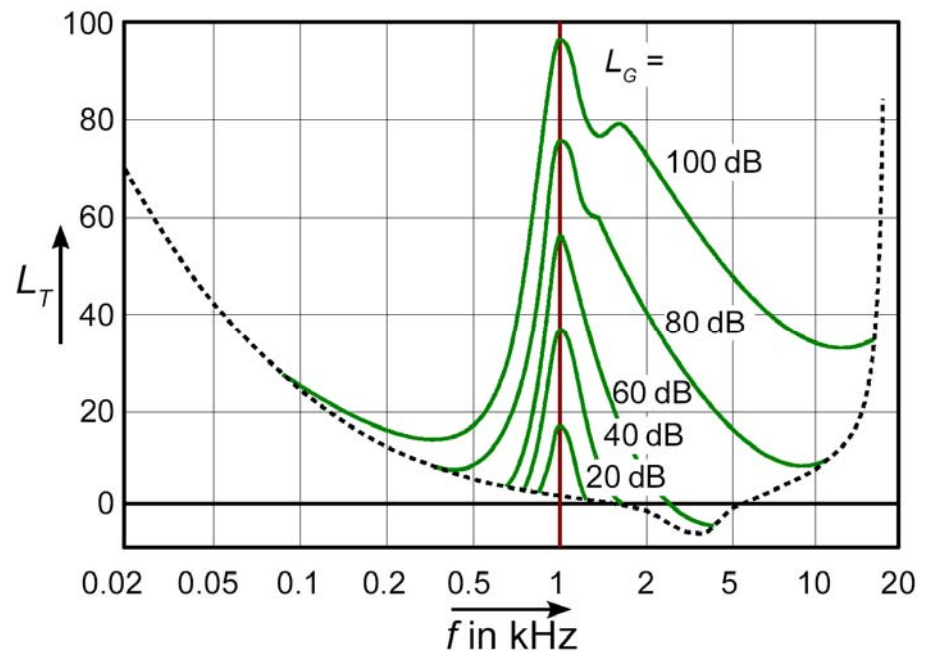
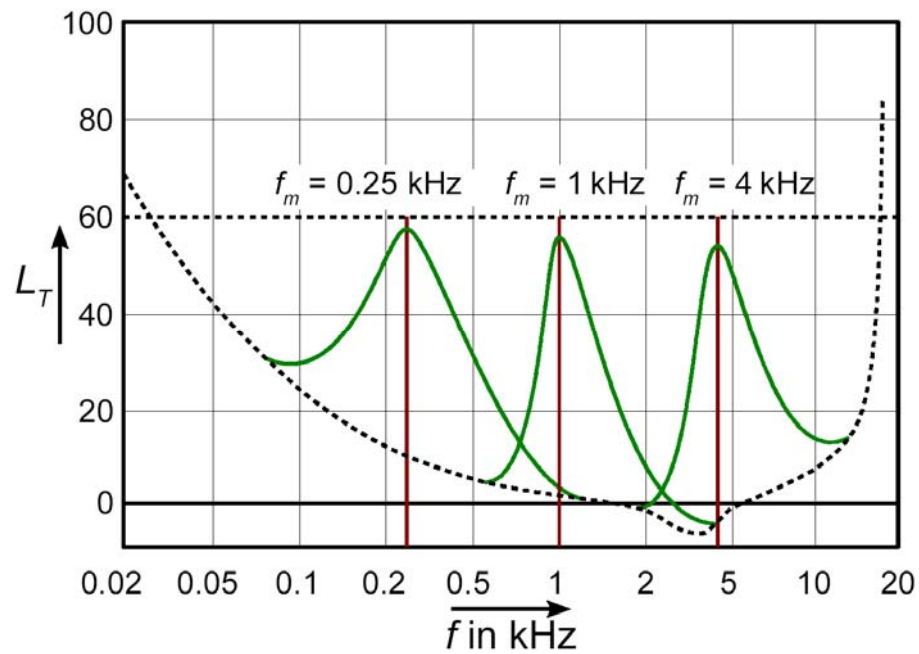
Filterkurven



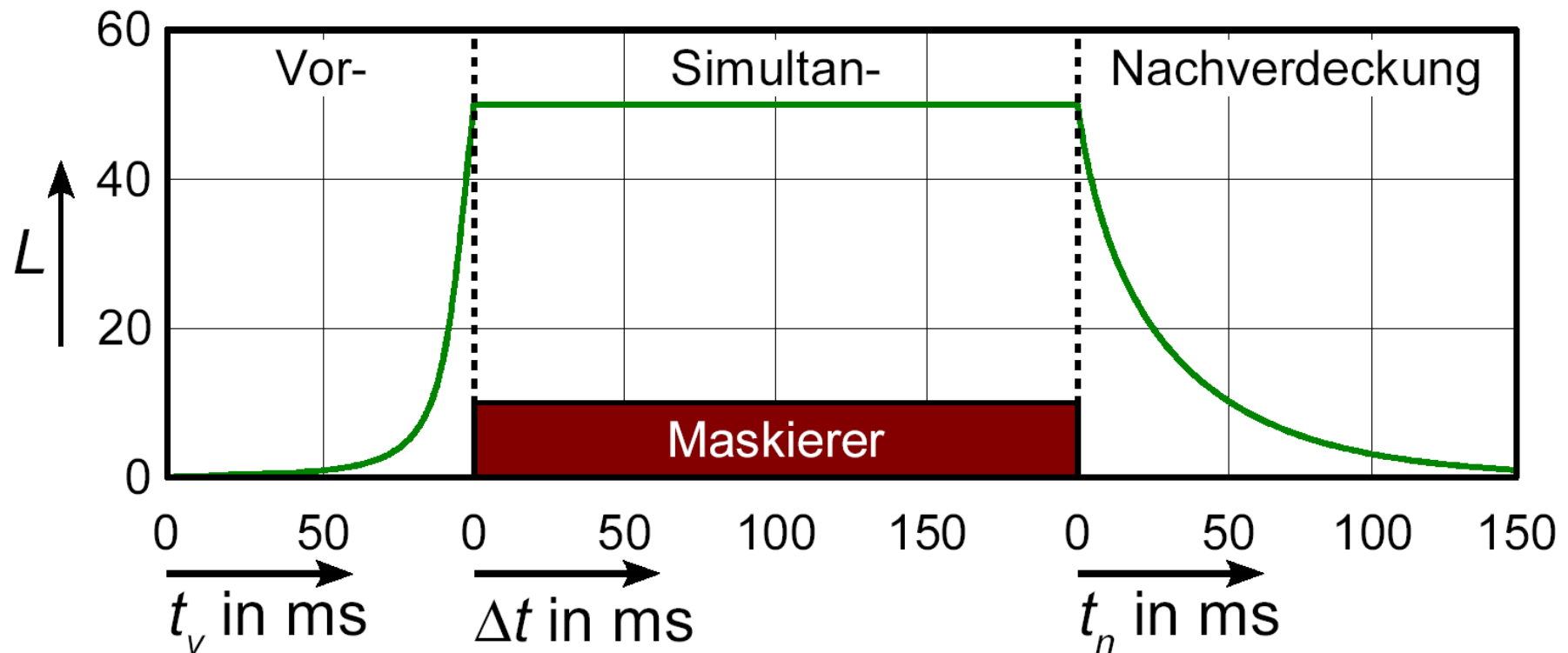
Frequenzgruppen, Barkskala, Länge der Basilarmembran



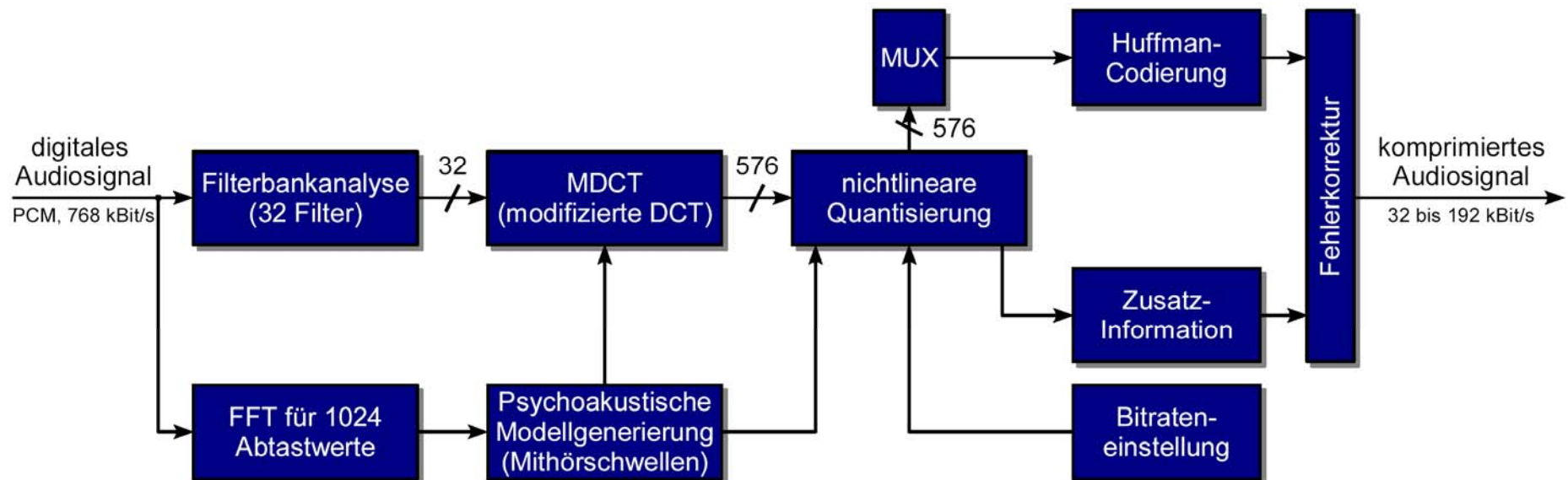
Spektrale Verdeckung



Zeitliche Verdeckung



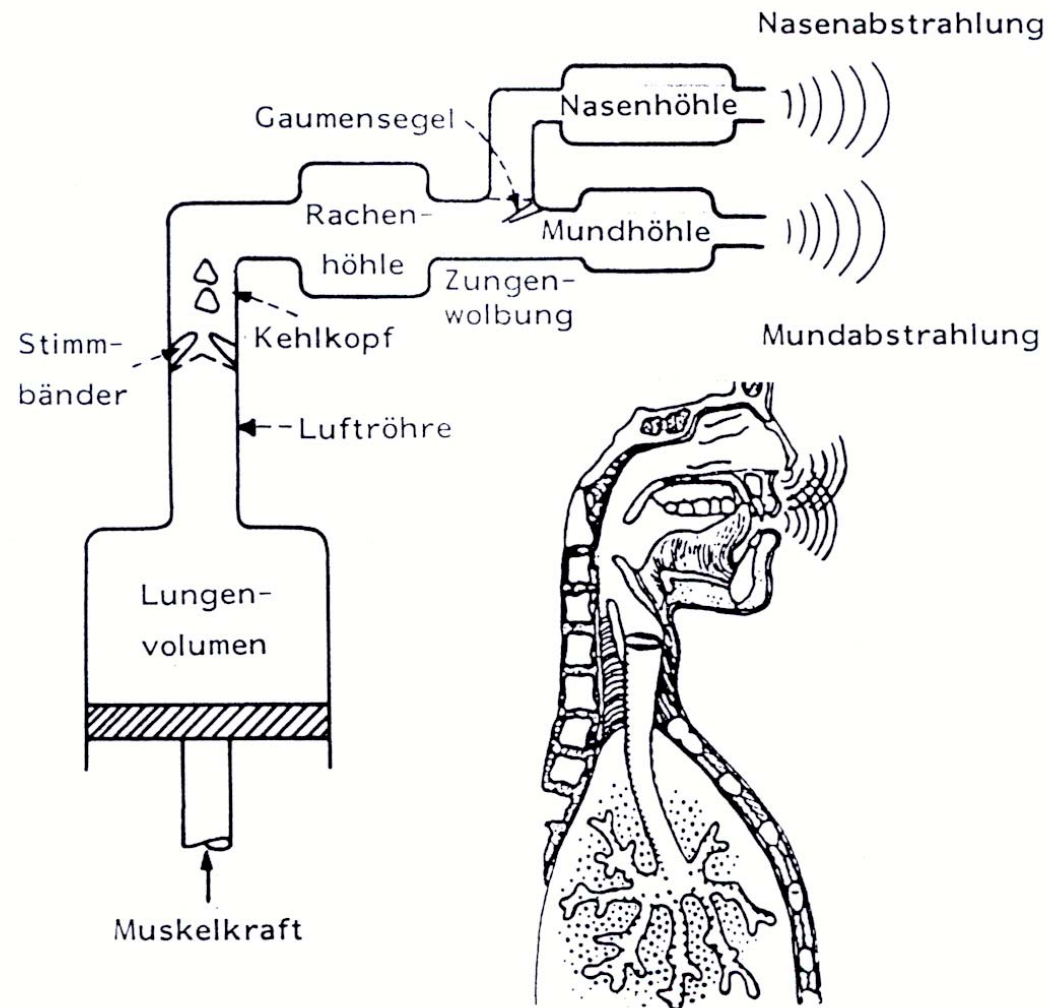
MP3 Codierungsschema



MP3 Dekodierungsschema



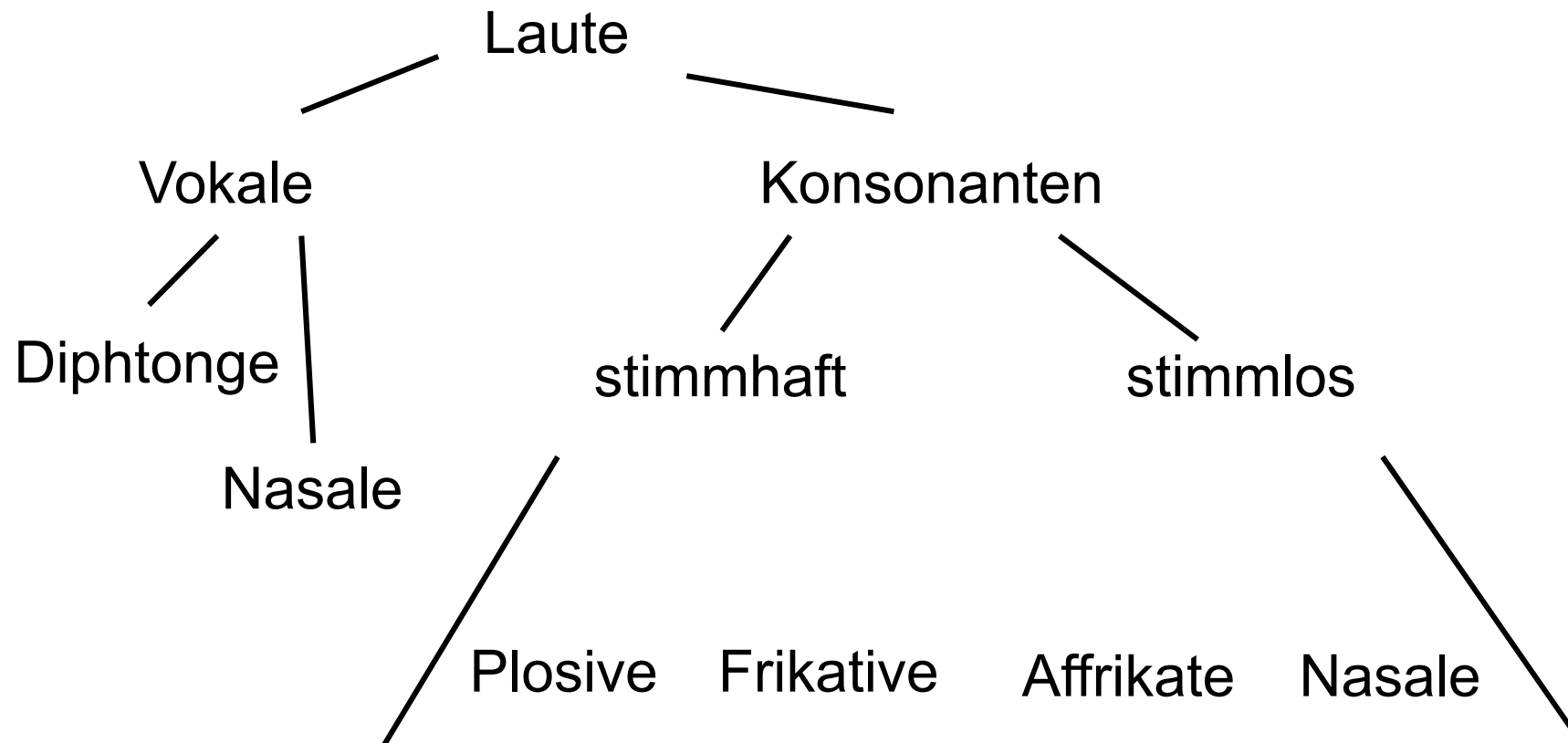
der menschliche Spracherzeugungsapparat



Phoneme: Aufzählung

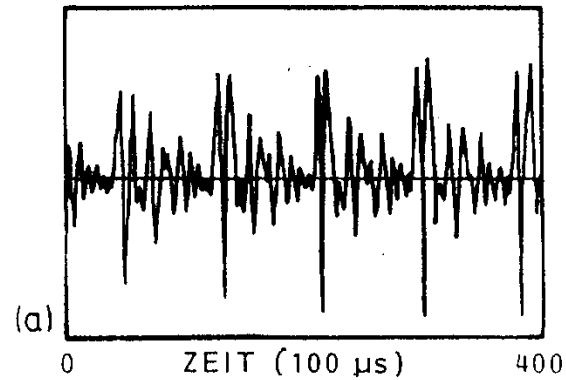
Phonem	Aussprache	Phonem	Aussprache	Phonem	Aussprache
/a/	K <u>a</u> mpf	/a:/	K <u>a</u> hn	/ai/	we <u>i</u> t
/au/	H <u>a</u> us	/ax/	ma <u>c</u> he	/b/	B <u>a</u> ll
/d/	d <u>e</u> utsch	/eh/	w <u>e</u> nn	/eh:/	Aff <u>a</u> re
/ey/	w <u>e</u> n	/f/	f <u>e</u> rn	/g/	g <u>e</u> rn
/h/	H <u>a</u> nd	/i/	H <u>i</u> mmel	/i:/	H <u>i</u> er
/j/	J <u>u</u> nge	/jh/	J <u>o</u> ystick	/k/	K <u>i</u> nd
/l/	l <u>i</u> nk	/m/	m <u>a</u> tt	/n/	N <u>e</u> st
/ng/	l <u>a</u> ng	/o/	o <u>ff</u> en	/o:/	O <u>ff</u> en
/oe/	H <u>o</u> lle	/oe:/	H <u>o</u> hle	/oy/	fre <u>u</u> t
/p/	P <u>a</u> ar	/r:/	r <u>e</u> nnen	/s/	f <u>a</u> ssen
/sh/	s <u>c</u> hön	/t/	T <u>a</u> fel	/u/	M <u>u</u> ter
/u:/	M <u>u</u> t	/v/	w <u>e</u> r	/x/	l <u>a</u> chen
/y/	T <u>y</u> p	/y:/	K <u>u</u> bel	/z/	s <u>i</u> ngen
/zh/	In <u>g</u> enieur	/sp/	„short pause“	/sil/	„silence“

Phoneme: Systematisch

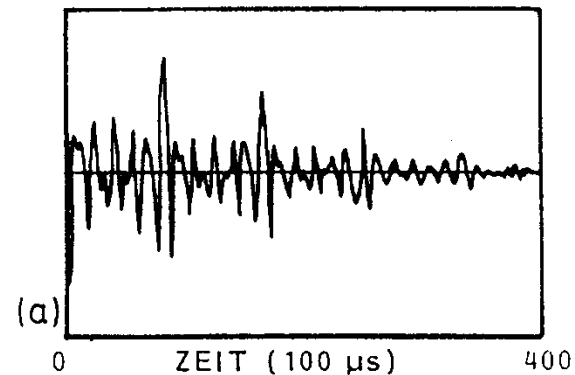
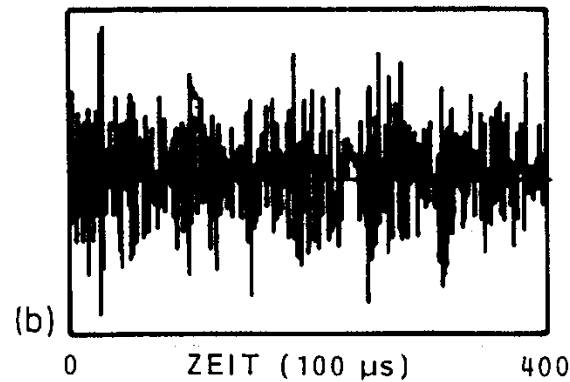


Signalverlauf und Korrelationsfunktion

SPRACHSIGNAL
A-LAUT AUS "NAME"



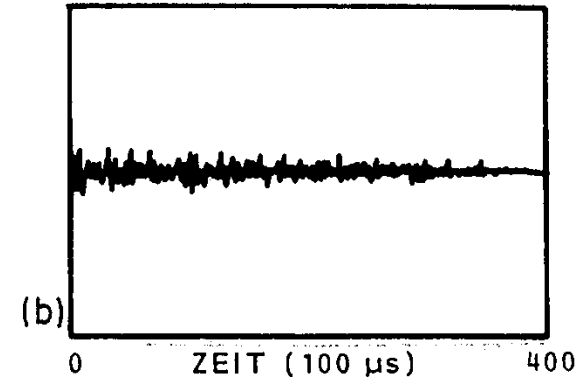
SPRACHSIGNAL
STIMMLOSER F-LAUT



AUTOKORRELATIONSFUNKTION,
RECHTECK-FENSTER

a = Stimmhaft

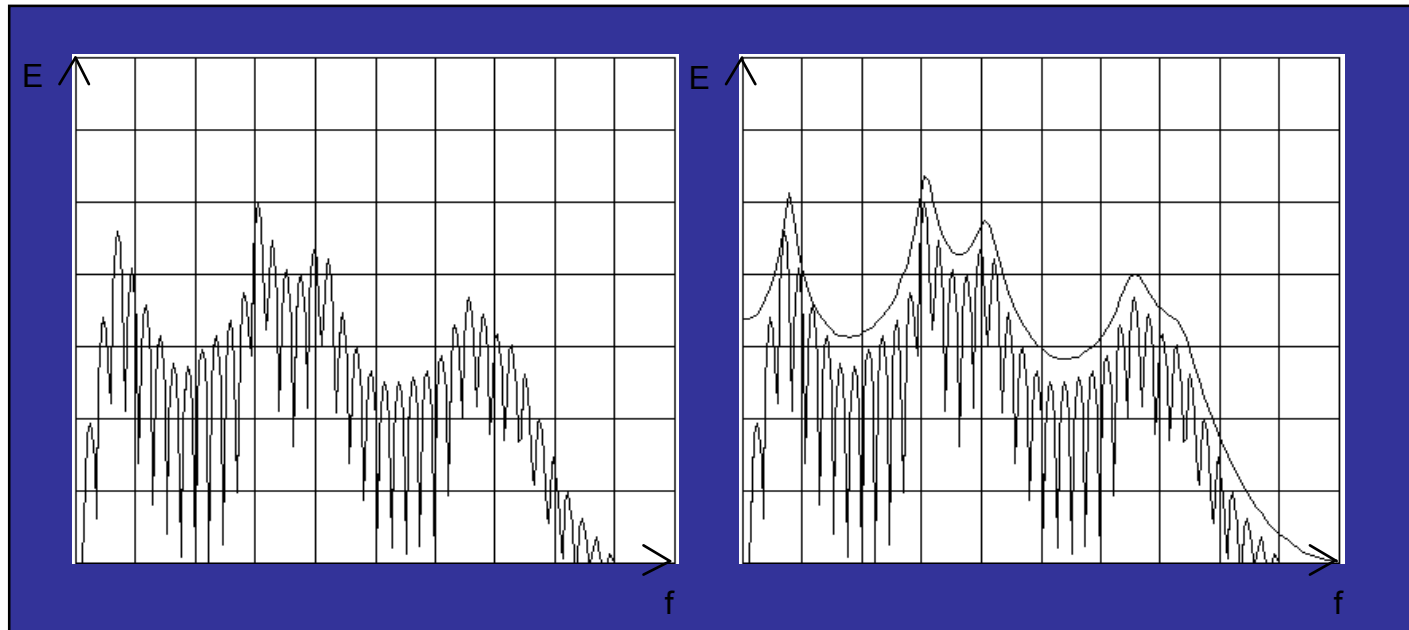
b = Stimmlos



AUTOKORRELATIONSFUNKTION,
RECHTECK-FENSTER

Quelle : Sichert

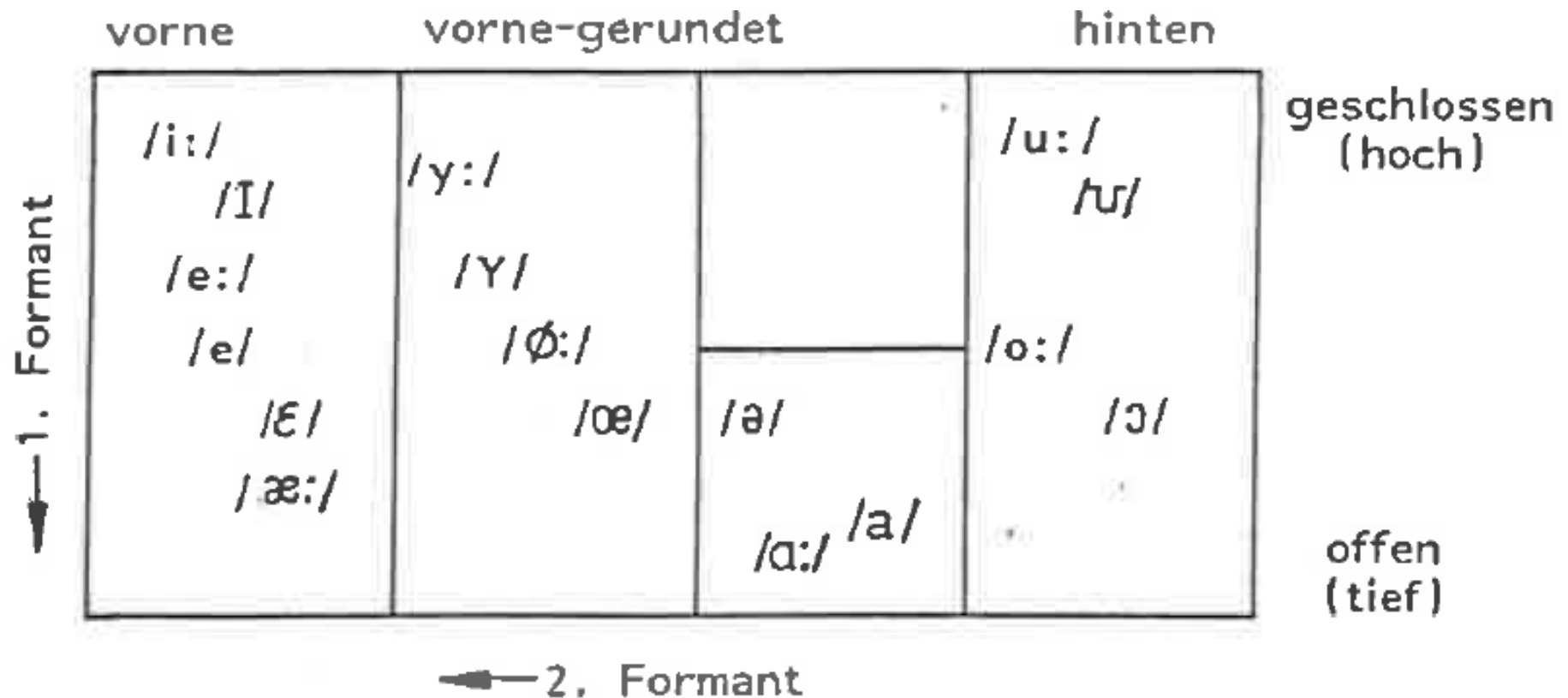
Ergebnis der Fourier-Transformation



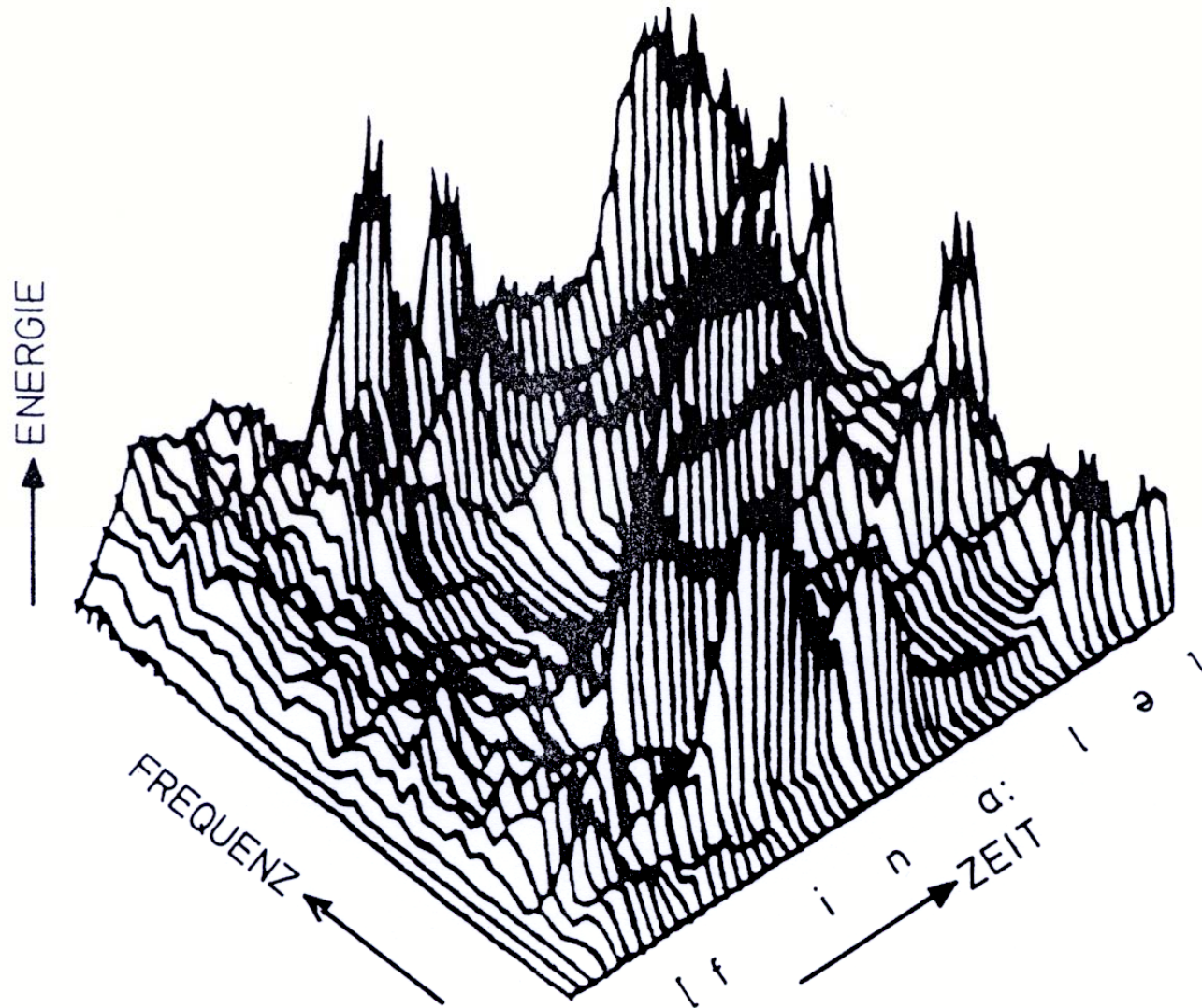
Originalergebnis

Ergebnis mit Hüllkurve

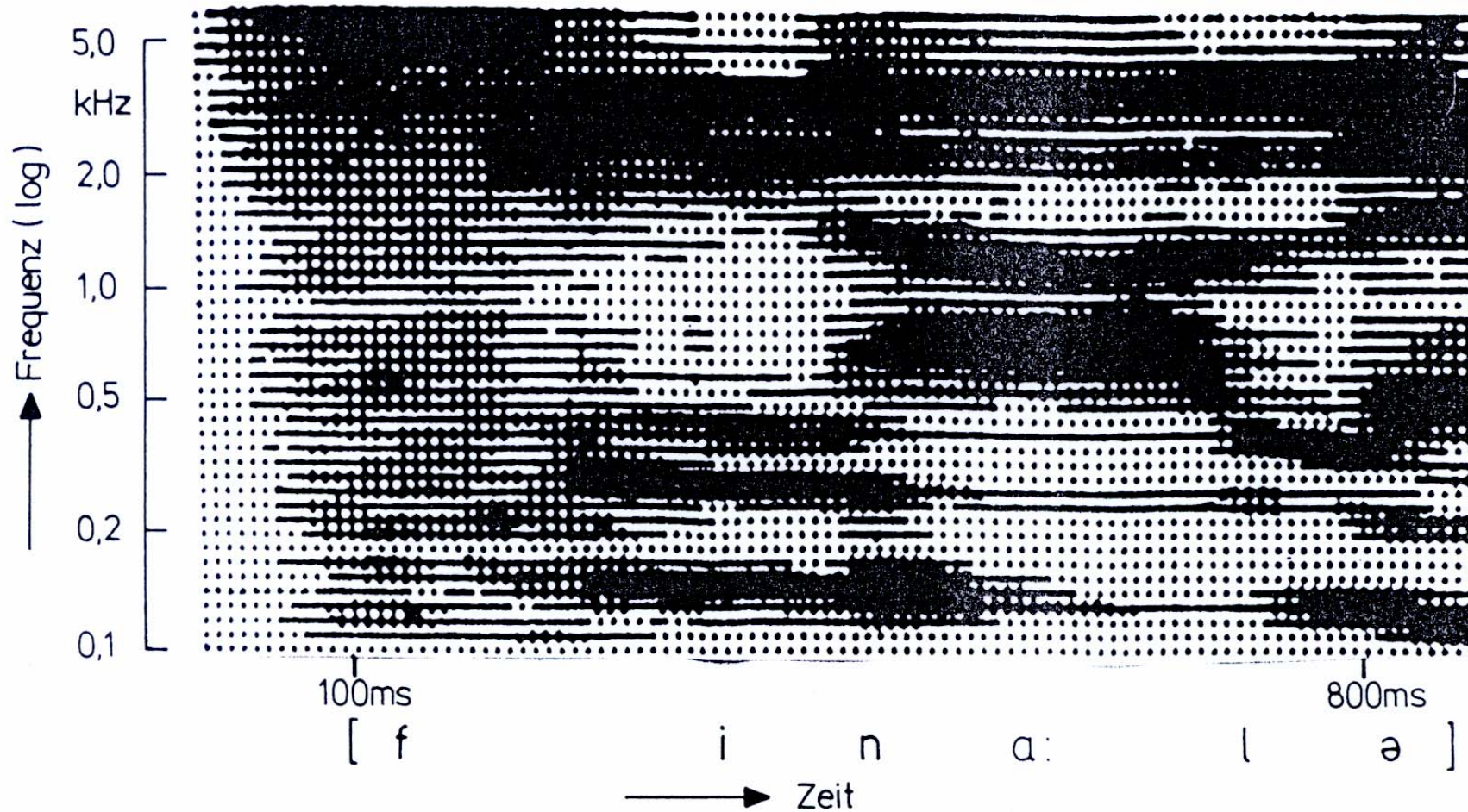
Formantkarte von Vokalen



3D-Darstellung eines Spektrogramms



2D-Darstellung eines Spektrogramms



Grammatik für natürliche Sätze

$\langle \text{Satz} \rangle \longrightarrow \langle \text{Hauptsatz} \rangle \quad \langle \text{Nebensatz} \rangle$
 $\langle \text{Hauptsatz} \rangle \longrightarrow \langle \text{Artikel} \rangle \quad \langle \text{Hauptsatz} \rangle$
 $\langle \text{Hauptsatz} \rangle \longrightarrow \langle \text{Nomen} \rangle$
 $\langle \text{Nomen} \rangle \longrightarrow \text{Sonne}$
 $\langle \text{Artikel} \rangle \longrightarrow \text{Die}$
 $\langle \text{Nebensatz} \rangle \longrightarrow \langle \text{Verb} \rangle$
 $\langle \text{Verb} \rangle \longrightarrow \text{scheint}$

$\langle \text{Satz} \rangle \longrightarrow \langle \text{Artikel} \rangle \quad \langle \text{Hauptsatz} \rangle \quad \langle \text{Nebensatz} \rangle$
 $\langle \text{Satz} \rangle \longrightarrow \langle \text{Artikel} \rangle \quad \langle \text{Nomen} \rangle \quad \langle \text{Nebensatz} \rangle$
 $\langle \text{Satz} \rangle \longrightarrow \langle \text{Artikel} \rangle \quad \langle \text{Nomen} \rangle \quad \langle \text{Verb} \rangle$
 $\langle \text{Satz} \rangle \longrightarrow \text{Die Sonne scheint}$

Generierung mathematischer Formeln

$\langle \text{Ausdruck} \rangle \longrightarrow \langle \text{Ausdruck} \rangle + \langle \text{Ausdruck} \rangle$

$\langle \text{Ausdruck} \rangle \longrightarrow \langle \text{Ausdruck} \rangle \cdot \langle \text{Ausdruck} \rangle$

$\langle \text{Ausdruck} \rangle \longrightarrow (\langle \text{Ausdruck} \rangle)$

$\langle \text{Ausdruck} \rangle \longrightarrow x$

$\langle \text{Ausdruck} \rangle \longrightarrow \langle \text{Ausdruck} \rangle \cdot \langle \text{Ausdruck} \rangle$

$\longrightarrow (\langle \text{Ausdruck} \rangle) \cdot \langle \text{Ausdruck} \rangle$

$\longrightarrow (\langle \text{Ausdruck} \rangle + \langle \text{Ausdruck} \rangle) \cdot \langle \text{Ausdruck} \rangle$

$\longrightarrow (\langle \text{Ausdruck} \rangle \cdot \langle \text{Ausdruck} \rangle + \langle \text{Ausdruck} \rangle) \cdot$
 $\quad \cdot \langle \text{Ausdruck} \rangle$

$\longrightarrow (x \cdot x + x) \cdot x = (x^2 + x) \cdot x$

Kontextfreie Grammatik (CFG)

$$\mathcal{G} = \{V, T, P, S\}$$

$V \equiv$ Variable, z. B. $\langle \text{Ausdruck} \rangle$ (Großbuchstaben)

$T \equiv$ Terminale, z. B. „x“ (Kleinbuchstaben)

$P \equiv$ Produktionsregel, z. B. $\langle \text{Ausdruck} \rangle \longrightarrow \text{„x“}$

$S \equiv$ Startsymbol

$$A \longrightarrow \alpha, \quad A \in \{V\} \text{ und } \alpha \in \{V \cup T\}$$

$$A \longrightarrow \alpha$$

$$A \longrightarrow \beta \quad A \longrightarrow \alpha \mid \beta \mid \gamma$$

$$A \longrightarrow \gamma$$

Chomsky-Normalform (CNF)

Die Standardform CNF enthält nur Produktionsregeln, bei denen auf der rechten Seite entweder nur zwei Variablen oder nur ein terminaler Ausdruck steht, also:

$A \rightarrow BC$ oder $A \rightarrow a$

Beispiel:

$S \rightarrow b A \mid a B$

$A \rightarrow b A A \mid a$

$B \rightarrow a B B \mid b$

1. Ersetzen von $S \rightarrow b A$ in $S \rightarrow C A$ und $C \rightarrow b$
2. Ersetzen von $S \rightarrow a B$ in $S \rightarrow D B$ und $D \rightarrow a$
3. Ersetzen von $A \rightarrow b A A$ in $A \rightarrow C A A$
4. Ersetzen von $A \rightarrow C A A$ in $A \rightarrow C E$ und $E \rightarrow A A$
5. Ersetzen von $B \rightarrow a B B$ in $B \rightarrow D B B$
6. Ersetzen von $B \rightarrow D B B$ in $B \rightarrow D F$ und $F \rightarrow B B$

Chomsky-Normalform (CNF)

Endergebnis der Transformation in CNF:

$$S \longrightarrow C A \mid D B$$

$$A \longrightarrow C E \mid a$$

$$B \longrightarrow D F \mid b$$

$$C \longrightarrow b$$

$$D \longrightarrow a$$

$$E \longrightarrow A A$$

$$F \longrightarrow B B$$

BNF und EBNF

$\langle \text{Ziffer außer Null} \rangle \longrightarrow 1 \mid 2 \mid 3 \mid 4 \mid 5 \mid 6 \mid 7 \mid 8 \mid 9$

$\langle \text{Ziffer} \rangle \longrightarrow 0 \mid \langle \text{Ziffer außer Null} \rangle$

$\langle \text{Ziffernfolge} \rangle \longrightarrow \langle \text{Ziffer} \rangle \langle \text{Ziffernfolge} \rangle \mid \langle \text{Ziffer} \rangle$

$\langle \text{positive ganze Zahl} \rangle \longrightarrow \langle \text{Ziffer außer Null} \rangle \mid \langle \text{Ziffer außer Null} \rangle \langle \text{Ziffernfolge} \rangle$

$\langle \text{Zahl} \rangle \longrightarrow \text{positive ganze Zahl} \mid - \langle \text{positive ganze Zahl} \rangle \mid 0$

$\text{ZifferAußerNull} \longrightarrow '1' \mid '2' \mid '3' \mid '4' \mid '5' \mid '6' \mid '7' \mid '8' \mid '9'$

$\text{Ziffer} \longrightarrow '0' \mid \text{ZifferAußerNull}$

$\text{Zahl} \longrightarrow [' -] \text{ZifferAußerNull} \{ \text{Ziffer} \} \mid '0'$

Kontextfreie Sprachen, einfache Grammatik

$$S \longrightarrow \underbrace{a S b}_{(1)} \mid \underbrace{a b}_{(2)}$$

Anwendung von (1): $S \longrightarrow a S b$
 $\longrightarrow a a S b b$
 $\longrightarrow a a a S b b b$
 \vdots
 $\longrightarrow a a \dots a S b b \dots b$

Anwendung von (2): $S \longrightarrow a \dots a a b b b \dots b = a^n b^n$

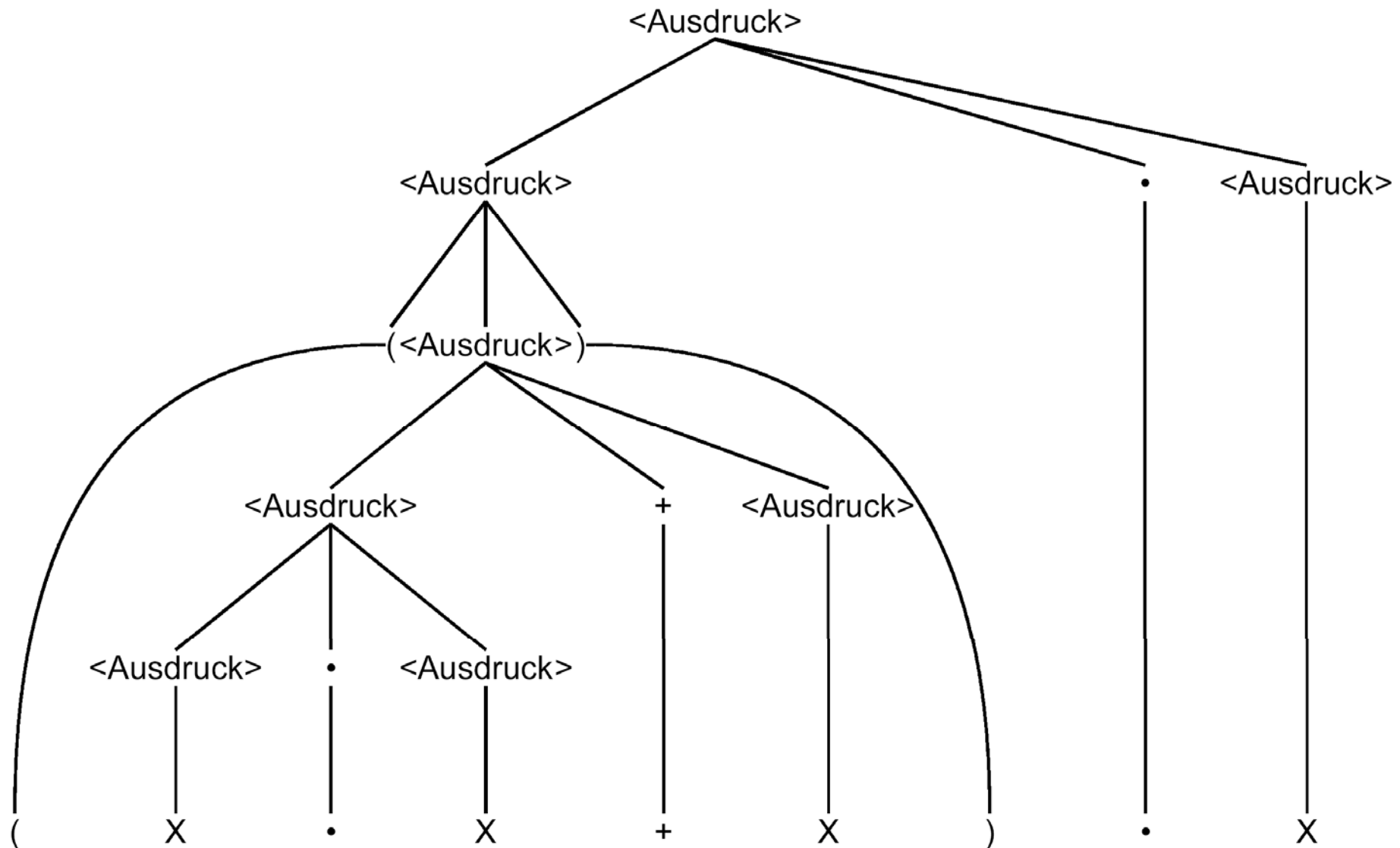
Komplexere Grammatik

$$\begin{aligned}
 S &\longrightarrow \underbrace{a B}_{(1)} \mid \underbrace{b A}_{(2)} \\
 A &\longrightarrow \underbrace{a S}_{(3)} \mid \underbrace{b A A}_{(4)} \mid \underbrace{a}_{(5)} \\
 B &\longrightarrow \underbrace{b S}_{(6)} \mid \underbrace{a B B}_{(7)} \mid \underbrace{b}_{(8)}
 \end{aligned}$$

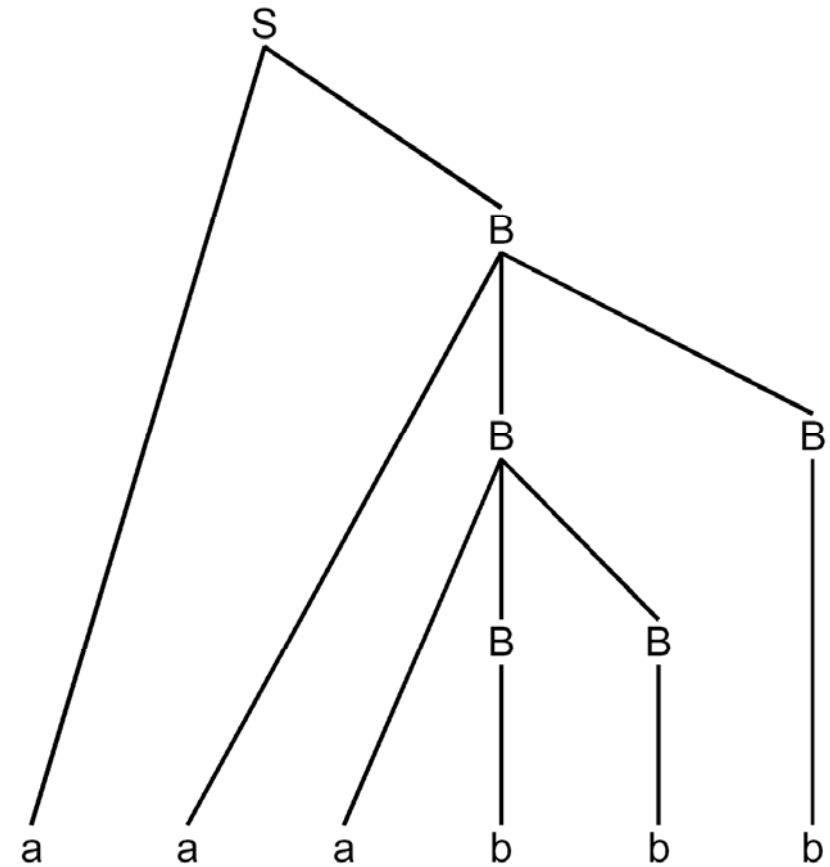
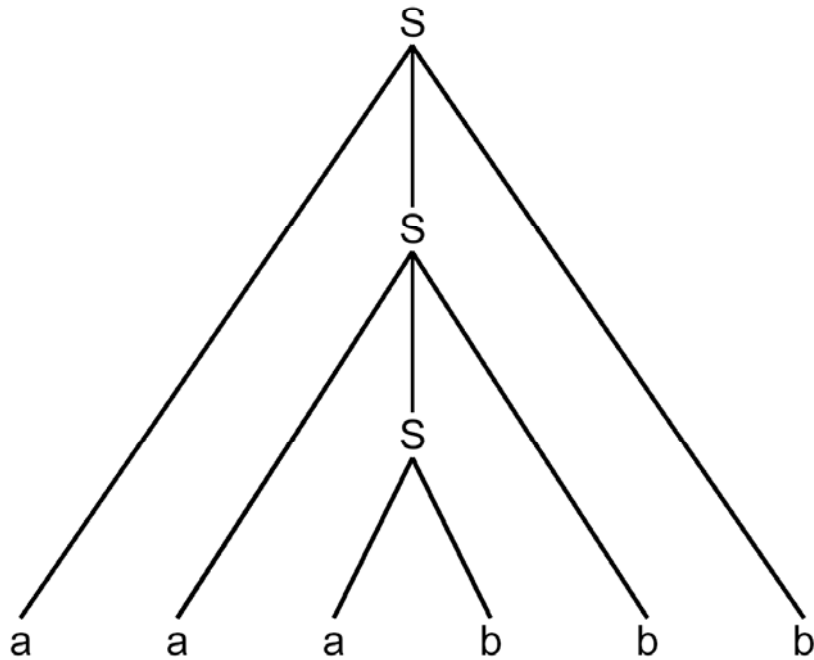
(Anzahl der Elemente a und b ist hier immer gleich)

Anwendung von (1)	$S \longrightarrow a B$
Anwendung von (7)	$\longrightarrow a a B B$
Anwendung von (8)	$\longrightarrow a a b b$
Anwendung von (2)	$S \longrightarrow b A$
Anwendung von (3)&(2)	$\longrightarrow b a b A$
Anwendung von (4)	$\longrightarrow b a b b A A$
	$\longrightarrow b a b b b A A b A A$
Anwendung von (5)	$\longrightarrow b a b b b a a b a a$

Parse-Tree für mathematische Formeln



Gleiche Symbolfolge geparkt mit verschiedenen Grammatiken



Ambiguitäten von Grammatiken:

Beispiel für uneindeutige Grammatik:

$$S \longrightarrow A B \mid B B$$

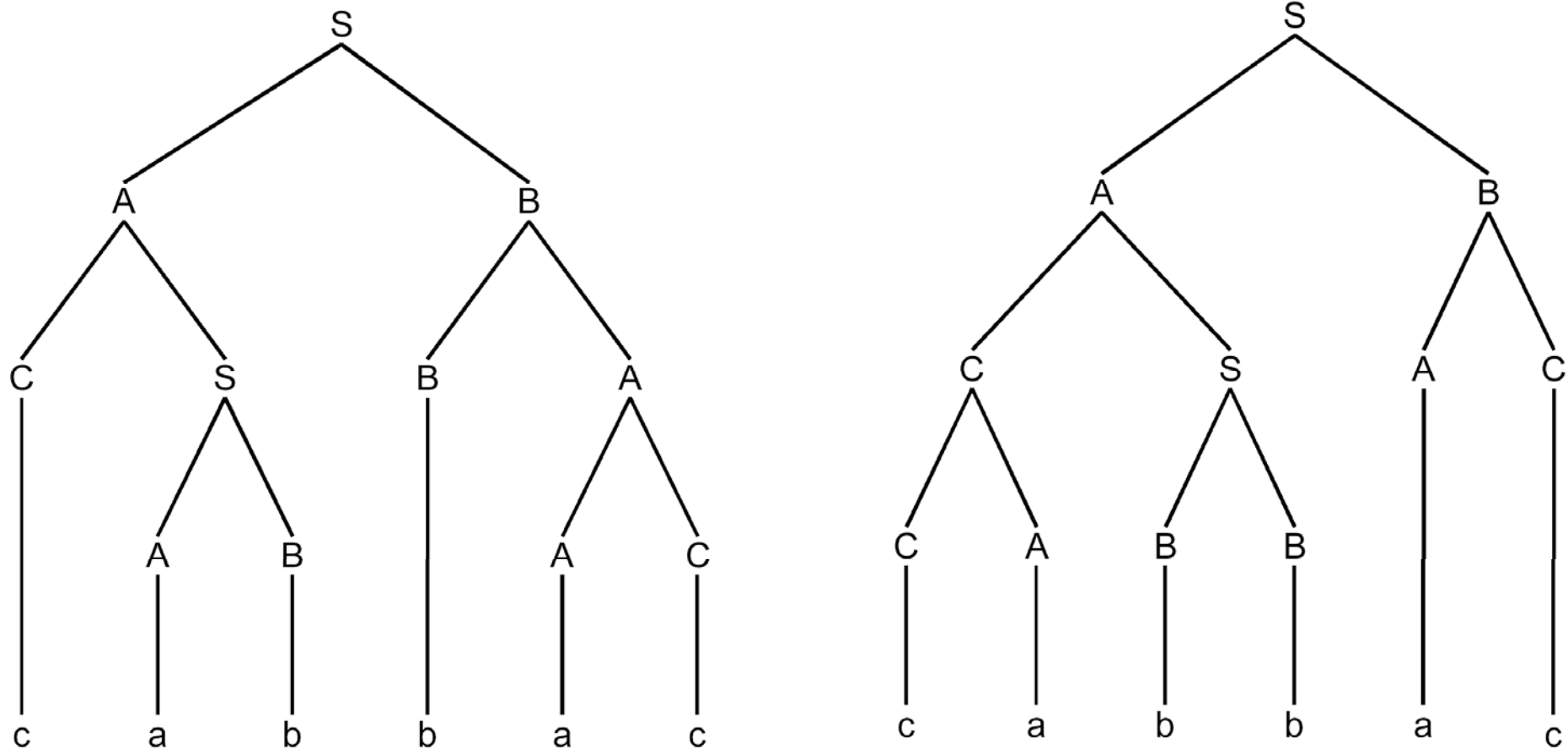
$$A \longrightarrow A C \mid C S \mid a$$

$$B \longrightarrow B A \mid A C \mid b$$

$$C \longrightarrow C A \mid c$$

$$S = (c, a, b, b, a, c)$$

Gleiche Symbolfolge, gleiche Grammatik, unterschiedlicher Parse-Tree:



Zerlegung natürlicher Sprache

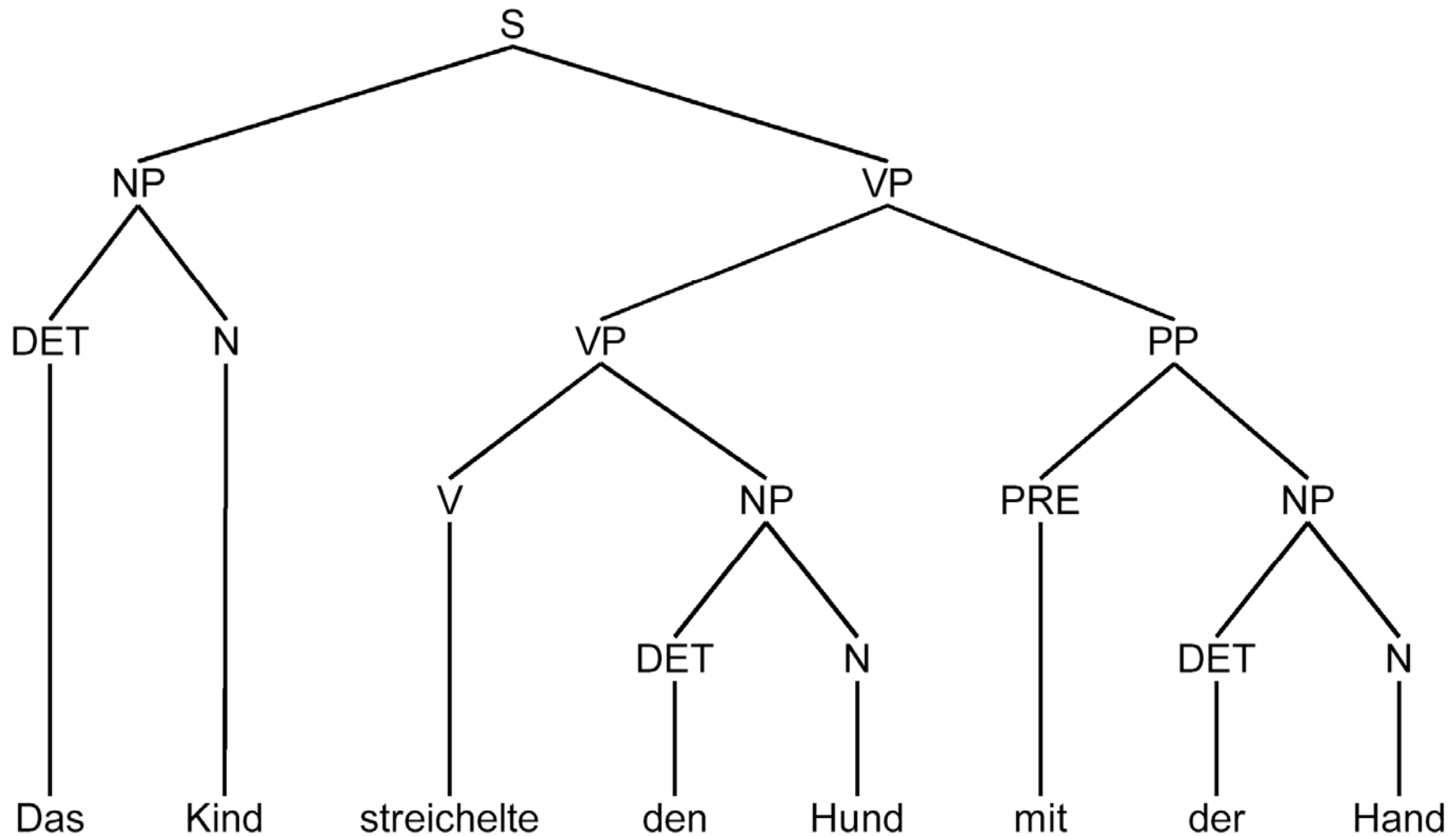
$$\mathcal{G} = \{V, T, P, S\}$$

$$V = \{NP, VP, AUX, ADJ, PRE, DET, V, N\}$$

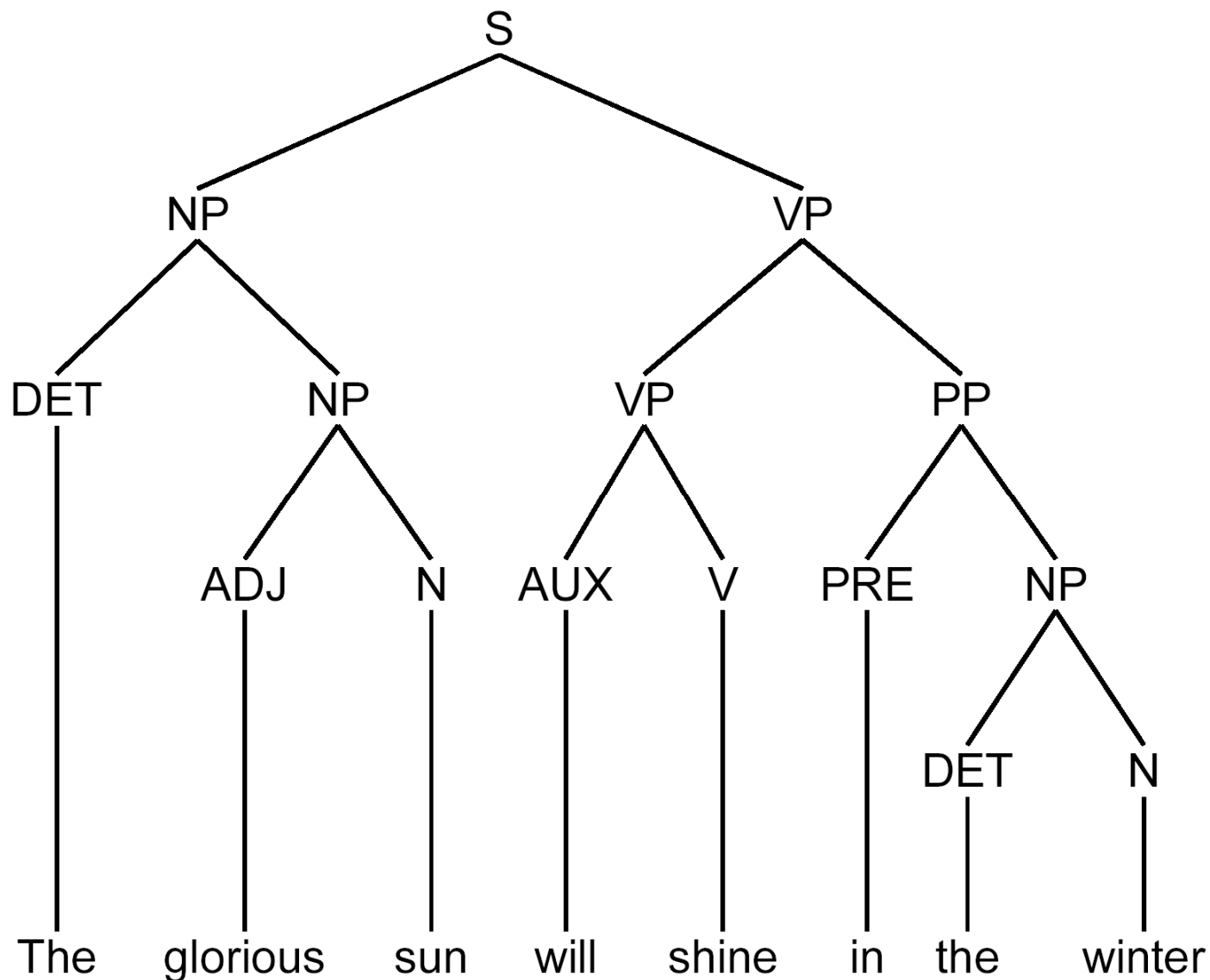
$$T = \{\text{der, die, das, } \dots, \text{groß, klein, } \dots, \text{wird, } \dots, \\ \text{streicheln, } \dots, \text{mit, in } \dots, \text{Junge, Hund, Hand, } \dots\}$$

$$P = \{ \begin{array}{ll} SA & \longrightarrow NP\ VP \mid VP\ NP, \\ NP & \longrightarrow DET\ N \mid ADJ\ N \mid DET\ NP \mid NP\ PP, \\ VP & \longrightarrow V\ NP \mid AUX\ V \mid V\ PP \mid V\ NP \mid VP\ PP \mid AUX\ VP, \\ PP & \longrightarrow PRE\ NP, \\ DET & \longrightarrow \text{„der“, „die“, „das“, } \dots, \\ ADJ & \longrightarrow \text{„klein“, „groß“, } \dots, \\ AUX & \longrightarrow \text{„wird“, } \dots, \\ V & \longrightarrow \text{„streicheln“, } \dots, \\ PRE & \longrightarrow \text{„in“, „mit“, } \dots, \\ N & \longrightarrow \text{„Junge“, „Hund“, „Hand“ } \dots \end{array} \}$$

Zerlegung natürlicher Sprache: Deutsch



Zerlegung natürlicher Sprache: Englisch








Geometrische Objekte als Grammatik

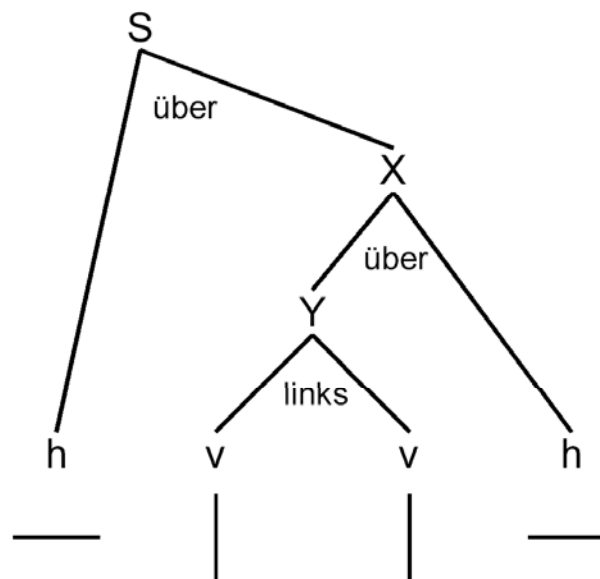
Prädikate: $\text{über}(X, Y) \quad \equiv \quad X \text{ ist über } Y$
 $\text{links}(X, Y) \quad \equiv \quad X \text{ ist links von } Y$

Grammatik: $S \longrightarrow \text{über}(h, X)$
 $X \longrightarrow \text{über}(Y, h)$
 $Y \longrightarrow \text{links}(v, v)$

Parse-Tree: Horizontale und vertikale Linien

vertikale Linie:  (v)
 horizontale Linie:  (h)

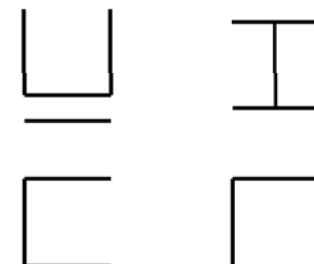
S: 
 X: 
 Y: 



erfolgreich geparste Objekte



zurückgewiesene Objekte



Definition eines Zustandsautomaten

Formal: $Z = (\mathcal{S}, \mathcal{X}, \mathbf{T}, s_0, \mathcal{F})$

\mathcal{S} ein Set mit einer endlichen Anzahl von Zuständen

\mathcal{X} zulässiges Alphabet für die zu verarbeitende Symbolfolge X

\mathbf{T} Transitionsfunktionen für die Zustände in \mathcal{S}

s_0 Anfangszustand

\mathcal{F} ein Set von festgelegten Endzuständen

Transitionsfunktion als Regel:

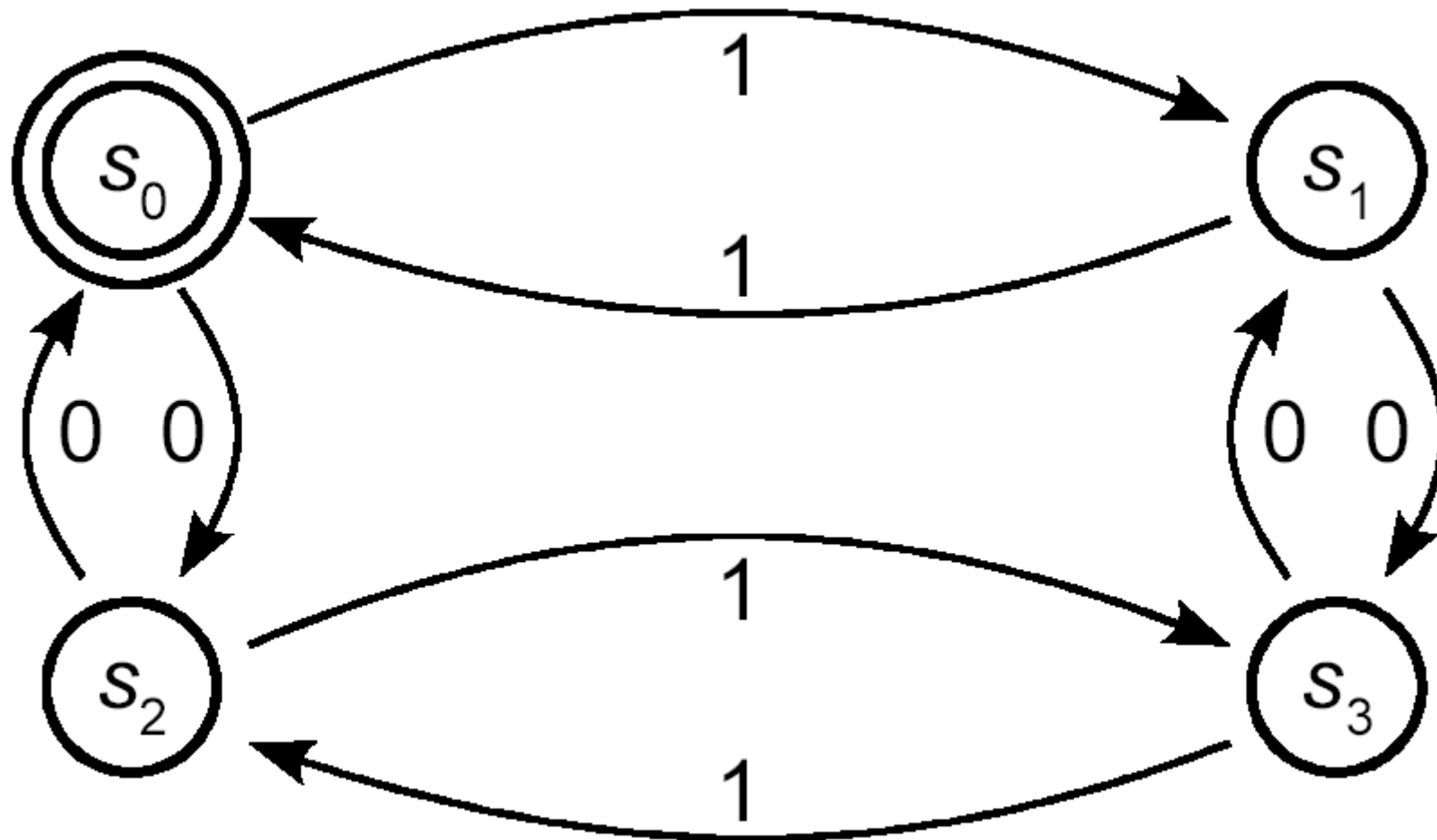
$$t(s^-, x_i) = s^+,$$

Beispiel: $\mathcal{S} = \{s_0, s_1, s_2, s_3\}$

$$\mathcal{X} = \{0, 1\}$$

$$\mathcal{F} = \{s_0\}$$

Zustandsautomat



Transitionsregeln in Tabellenform

alter Zustand	Symbol-Input	
	0	1
s^-		
s_0	s_2	s_1
s_1	s_3	s_0
s_2	s_0	s_3
s_3	s_1	s_2

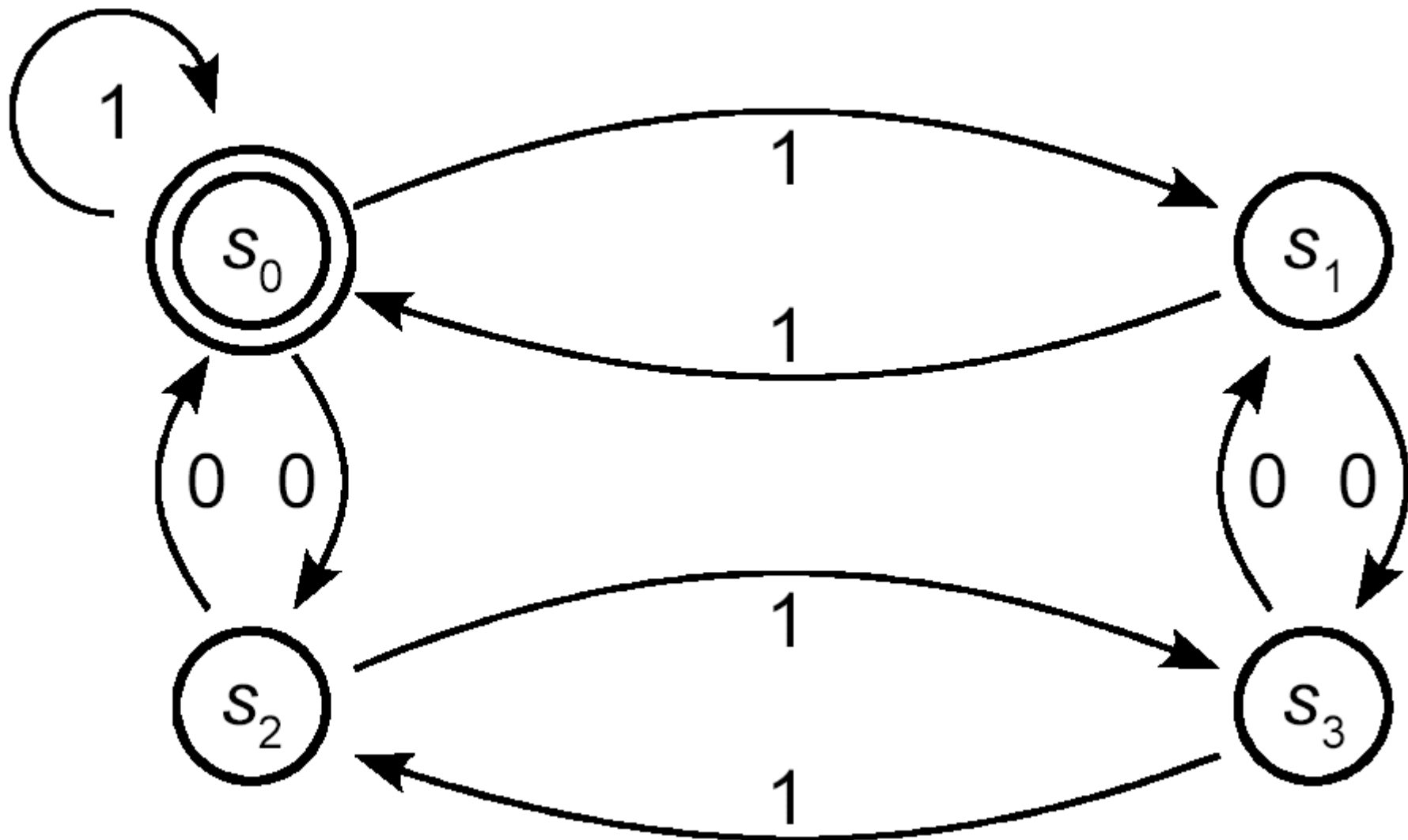
Verarbeitung einer Symbolfolge

Symbolfolge	1	0	1	1	0	
Zustandsfolge	s_0	s_1	s_3	s_2	s_3	s_1

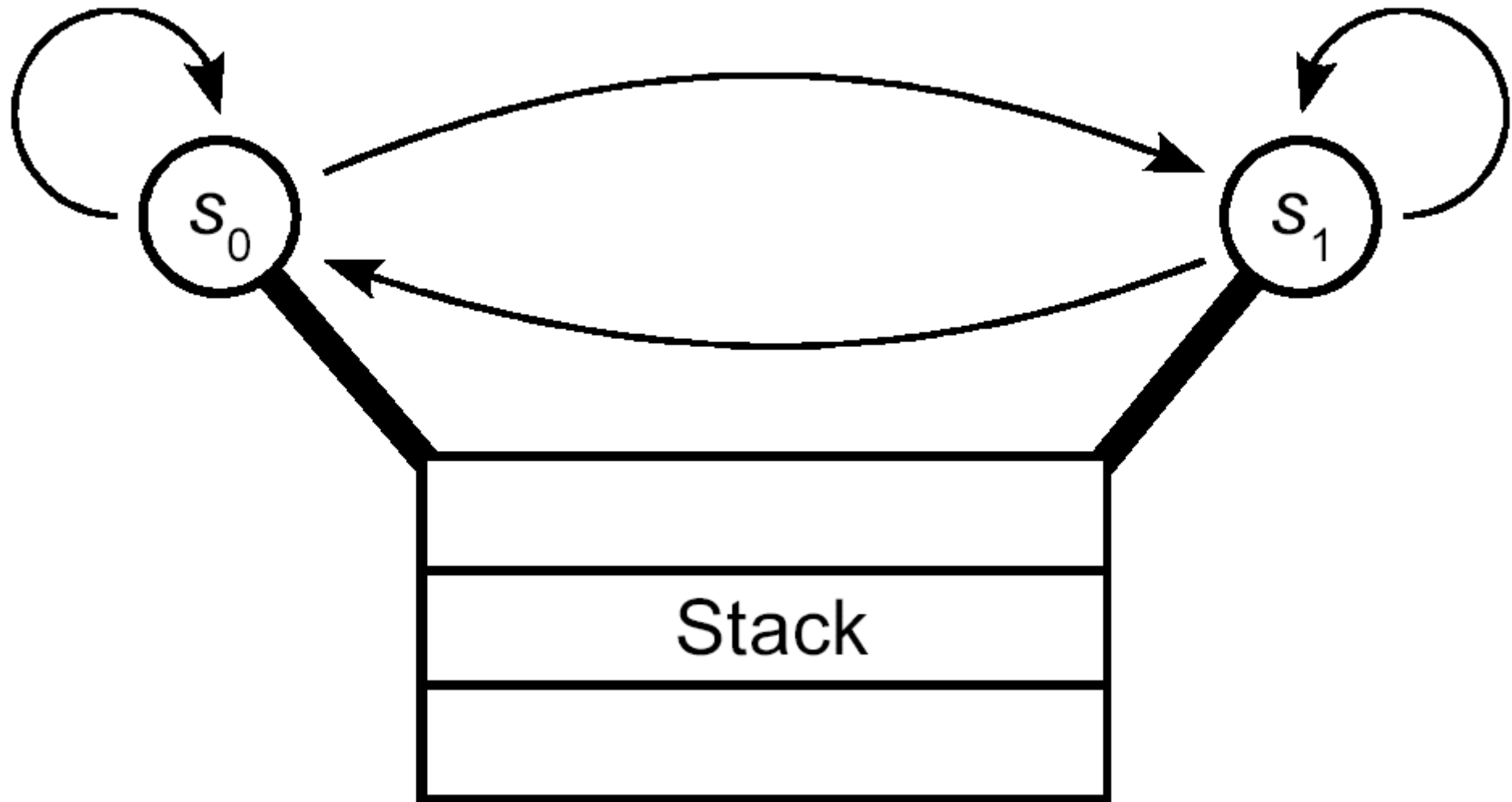
Hinzufügen des Symbols '1' würde zur Akzeptanz der Gesamtfolge führen

(Anzahl der Symbole '0' und '1' muss jeweils gerade sein)

Nicht-deterministischer Zustandsautomat



Kellerautomat (pushdown automata)



Formale Definition eines Kellerautomaten

$$K = (\mathcal{S}, \mathcal{X}, \mathcal{Y}, \mathbf{T}, s_0, y_0, \mathcal{F})$$

\mathcal{S} ein Set mit endlichen Anzahl von Zuständen

\mathcal{X} zulässiges Alphabet für die zu verarbeitende Symbolfolge X

\mathcal{Y} zulässiges Alphabet für die Symbole im Stack

\mathbf{T} Transitionsfunktionen für die Zustände in \mathcal{S}

s_0 Anfangszustand

y_0 Startsymbol für den Stack

\mathcal{F} ein Set von festgelegten Endzuständen (leere Menge für den Fall, dass Akzeptanz von X über den leeren Stack definiert ist)

Beispiel für einen Kellerautomaten

$$S \rightarrow 0 S 0 \mid 1 S 1 \mid z$$

Grammatik:

Erzeugt asymmetrische Symbolfolgen, z.B. der Form: $X = (0\ 0\ 1\ 0\ 1\ z\ 1\ 0\ 1\ 0\ 0)$

Kellerautomat zur Verarbeitung dieser Folgen:

$$\mathcal{S} = \{s_0, s_1\}$$

$$\mathcal{X} = \{0, 1, z\}$$

$$\mathcal{Y} = \{a, b, c\}$$

(hier entspricht a dem Symbol 0, b dem Symbol 1
und c dem Stack" Startsymbol.)

$$y_0 = c$$

Transitionsregeln für diesen Kellerautomaten

oberstes Stacksymbol	alter Zustand s^-	Symbol-Input		
		0	1	,z'
,a'	s_0	Verbleib in s_0 , ,a' auf Stack	Verbleib in s_0 , ,b' auf Stack	Trans zu s_1 , Stack bleibt
	s_1	Verbleib in s_1 , reduziere Stack	—	—
,b'	s_0	Verbleib in s_0 , ,a' auf Stack	Verbleib in s_0 , ,b' auf Stack	Trans zu s_1 , Stack bleibt
	s_1	—	Verbleib in s_1 , reduziere Stack	—
,c'	s_0	Verbleib in s_0 , ,a' auf Stack	Verbleib in s_0 , ,b' auf Stack	Trans zu s_1 , Stack bleibt
	s_1	entferne sofort oberstes Stacksymbol		

Kellerautomat für die Verarbeitung einer Symbolfolge

asymmetrische Symbolfolge, z.B.: $X = (0\ 0\ 1\ 0\ z\ 0\ 1\ 0\ 0)$

Verarbeitung durch Kellerautomaten:

Kellerautomat zur Verarbeitung einfacher Symbolfolgen

Symbolfolge: $X = (0^n 1^n)$

Transitionsformel der Form (y, x, op)

