

P4.1 Einführung in die Signalverarbeitung

Phonetik und Sprachverarbeitung, 2. Fachsemester,
Block 4 'Sprachtechnologie I'

Florian Schiel

Institut für Phonetik und Sprachverarbeitung, LMU München

Signalverarbeitung - Teil 1

Allgemeines

- Unterrichtssprache ist Deutsch (englische Fachbegriffe in Klammern)
- Fragen am besten sofort; besser einmal zuviel gefragt
- Literatur (* leicht — *** anspruchsvoll):
 - * **Jurafsky D, Martin J H (2000): Speech and Language Processing. Prentice Hall, Kap I.7.**
 - ** Schrüfer E (1980): Signalverarbeitung
 - ** **Pfister B, Kaufmann T (2008): Sprachverarbeitung - Grundlagen und Methoden der Sprachsynthese und Spracherkennung. Springer-Verlag Berlin Heidelberg.**
 - ** Steven W. Smith (1997): <http://www.dspguide.com/pdfbook.htm>
 - * Schnupp, Nelken & King (2011): Auditory Neuroscience. Chapter 1. MIT Press Cambridge MA USA.
 - *** Rabiner, Lawrence R., Schafer R W (1978): Digital Processing of Speech Signals. Prentice-Hall, New Jersey, USA.
 - *** Hess W (1993): Digitale Filter. Teubner Studienbücher, B.G.Teubner, Stuttgart.
 - * Harrington J, Cassidi St (1999): Techniques in Speech Acoustics. Kluwer Academic Publishers, Dordrecht/Boston/London.

Akustisches Sprachsignal

Sprachsignal = akustisches Signal = Schalldrucksignal

Schalldrucksignal (*sound pressure*) $p(t)$

= Änderung des Luftdrucks; Luftdruck $\approx 10^4 Pa$

Einheit *1 Pascal* = $1 N/m^2$

Bezugswert (*reference value*) $p_0 = 2 \cdot 10^{-5} Pa$

Bereich Schalldruck $10^{-5} - 10^3$ (8 Größenordnungen!)

Umrechnung in logarithmische Pegel (*level*):

$$L[dB] = 20 \cdot \log_{10} \frac{p_{eff}}{p_0}$$

Absoluter Pegel: Bezugspunkt $0dB$: p_0

Relativer Pegel: Bezugspunkt $0dB$ beim maximal messbaren Wert
(Werte $> 0dB$: übersteuert)

Faustregeln

+20dB : Faktor 10

-20dB : Faktor 1/10

+6dB : Faktor 2

-6dB : Faktor 1/2

Demo

Demo: Schallpegelmessgerät

Auf Bezugswert p_0 geeichtes Messgerät für Schallpegel

Demo: SpeechRecorder

Auf maximalen Wert 0db geeichtes Aufnahmegerät für Sprache

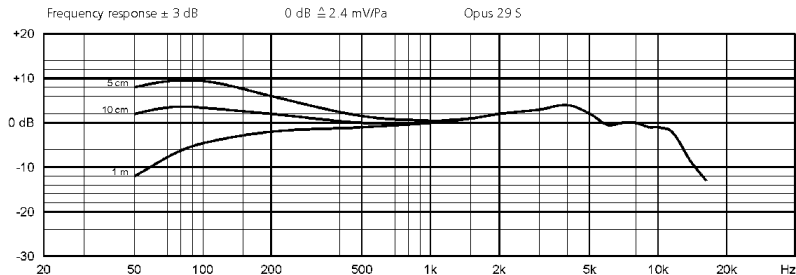
Mikrophone

Mikrofon wandelt

Schalldrucksignal $p(t)$ in Spannungssignal $u(t)$.

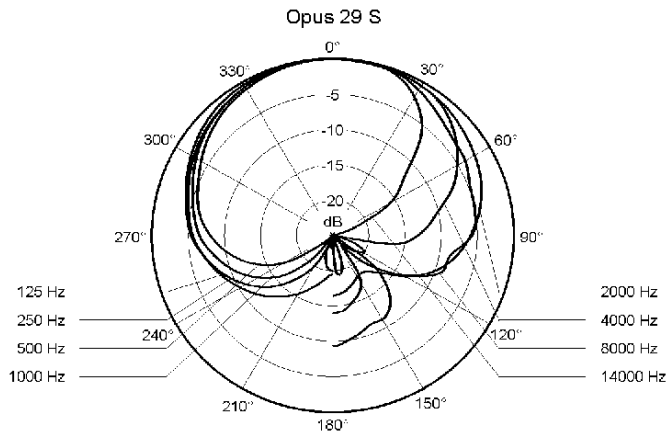
$$u(t) \approx p(t)$$

Frequenzgang (*frequency response*): verschiedene Frequenzen werden verschieden empfindlich in Spannung transformiert.



Mikrophone Richtwirkung

Richtwirkung (*directional effect*): Empfindlichkeit variiert mit der Richtung, meistens: Keulen- oder Nieren-Charakteristik.

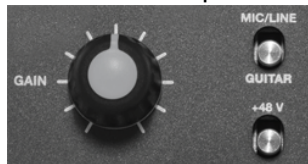


Mikrophone

Häufig: Elektret-Mikrophon (*condensor microphone*):
Leitfähige Membran bewegt sich in elektrischen Feld.
Dazu Vorspannung (Phantomspannung *phantom power*)
notwendig.

Standard ist 48 V.

Vorverstärker speist Gleichspannung im Signalweg ein.



Vorverstärker

Vorverstärker (*pre-amplifier*) ist immer notwendig vor Digitalisierung.

Auch hier: Frequenzgang beachten. Frequenzgang von Mikrophon und Vorverstärker werden multipliziert.

Brummstörung

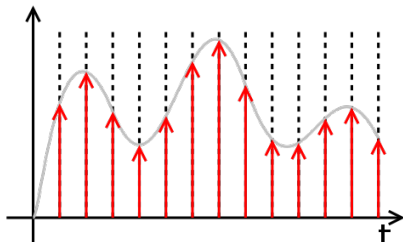
Durch schlechte (hoch-ohmige) Erdung wird in die *Masseleitung* des Mikrophons eine 50Hz-Wechselspannung induziert (Elektromagnetisches Signal verursacht durch Energienetz).

Besonders, wenn eine geschlossene Schleife aus Masseleitungen entsteht (*Brummschleife*).

Analog-Digital-Wandlung

Analog-Digital-Wandler (analog/digital converter, A/D):
Messung der Spannung in festem Zeitraster T_{abt}
Digitales Signal = Kette von Messwerten = Tabelle von Zahlen
Abtastrate (*sampling rate*) in Hz bestimmt Abstand der
Messpunkte:

$$f_{abt} = \frac{1}{T_{abt}} \quad \text{z.B.} \quad T_{abt} = 0.00005 \text{sec} \quad f_{abt} = 20 \text{kHz}$$

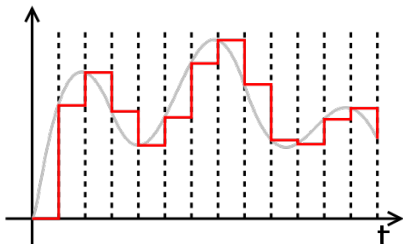


$$s(t) \rightarrow s(nT_{abt}), n = 0 \dots N$$

$$s(nT_{abt}) \equiv s(t_n)$$

Digital-Analog-Wandlung (Rekonstruktion)

Rekonstruktion des analogen Signals:
Treppenfunktion und anschließender Tiefpass (anti-aliasing filter) um Oberwellen der Treppenfunktion auszufiltern.



Zahlenformate (Kodierung)

Zahlenformat (coding) bestimmt Genauigkeit der Messung pro Abtastwert:

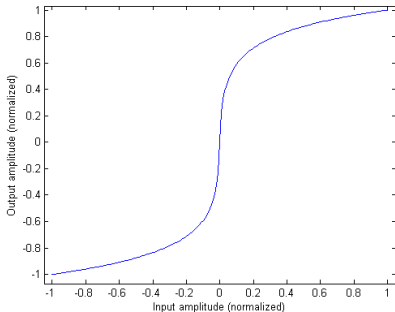
Meistens Integerzahlen mit 16bit : -32767 ... +32767

Auch gebräuchlich: 8bit (Telefon), 24bit (HiFi)

Bei Telefon: Nicht-lineare Kennlinie 8bit \rightarrow 12bit

Amerikanisch: μlaw , europäisch: $alaw$

μ -law Kennlinie
x-Achse: 12 bit
y-Achse: 8 bit



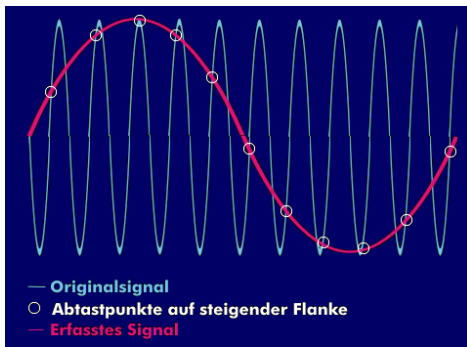
Abtasttheorem (Abtastbedingung)

Abtastbedingung (Shanon theorem, Nyquist-Bedingung)

Die Abtastrate f_{abt} muss mindestens doppelt so groß sein wie die maximale Frequenz f_{max} im Signal.

$$f_{abt} \geq 2f_{max}$$

Unterabtastung
(aliasing)



Abtasttheorem (Abtastbedingung)

Daher immer notwendig:

Analoger Tiefpassfilter vor der Abtastung!

Tiefpassfilter filtert alle Frequenzen oberhalb der Grenzfrequenz aus dem Signal.

Grenzfrequenz muss deutlich unter der halben Abtastrate liegen!

Abtasttheorem (Abtastbedingung)

Typische Werte für f_{abt} :

8kHz (Telefon) → Frequenzbereich 0...4000Hz

16kHz (Sprache) → Frequenzbereich 0...8000Hz

44100Hz (HiFi) → Frequenzbereich 0...22050Hz

Demo

Praat:

1000Hz Sinuston mit

16000, 5000, 2010, 1800Hz Abtastrate

Änderung der Abtastrate

Änderung der Abtastrate eines digitalen Signals (re-sampling)

Erhöhung (up-sampling):

immer möglich, neuer Frequenzbereich ist leer

Erniedrigung (down-sampling):

nur erlaubt, wenn sicher, dass Abtastbedingung erfüllt

Frequenzbereich wird kleiner

(ev. mit digitalem Tiefpass vorher filtern)

Berechnung der neuen Abtastwerte durch *Interpolation* oder
Rekonstruktion

Demo

*Praat : up-sampling eines Sprachsignals mit 16kHz Abtastrate
→ vergrößerter Frequenzbereich, aber dieser ist leer.*

Fragen

Luftdruck ist von der absoluten Höhe abhängig. Schall besteht aus kleinsten Luftdruckschwankungen. Warum 'hört' man dann nicht den Luftdruckunterschied, wenn man eine Treppe hinaufgeht?

Ein Signal ist +26dB lauter als ein anderes. In welchem Verhältnis stehen die Schalldrucke zueinander?

Ein Signal enthält als maximale Frequenz 10kHz. Wie groß ist der maximale Abstand der Abtastwerte bei der Digitalisierung?

Ein Signal wird mit 16bit abgetastet. Wie groß ist dann der theoretisch mögliche Pegelumfang in dB?