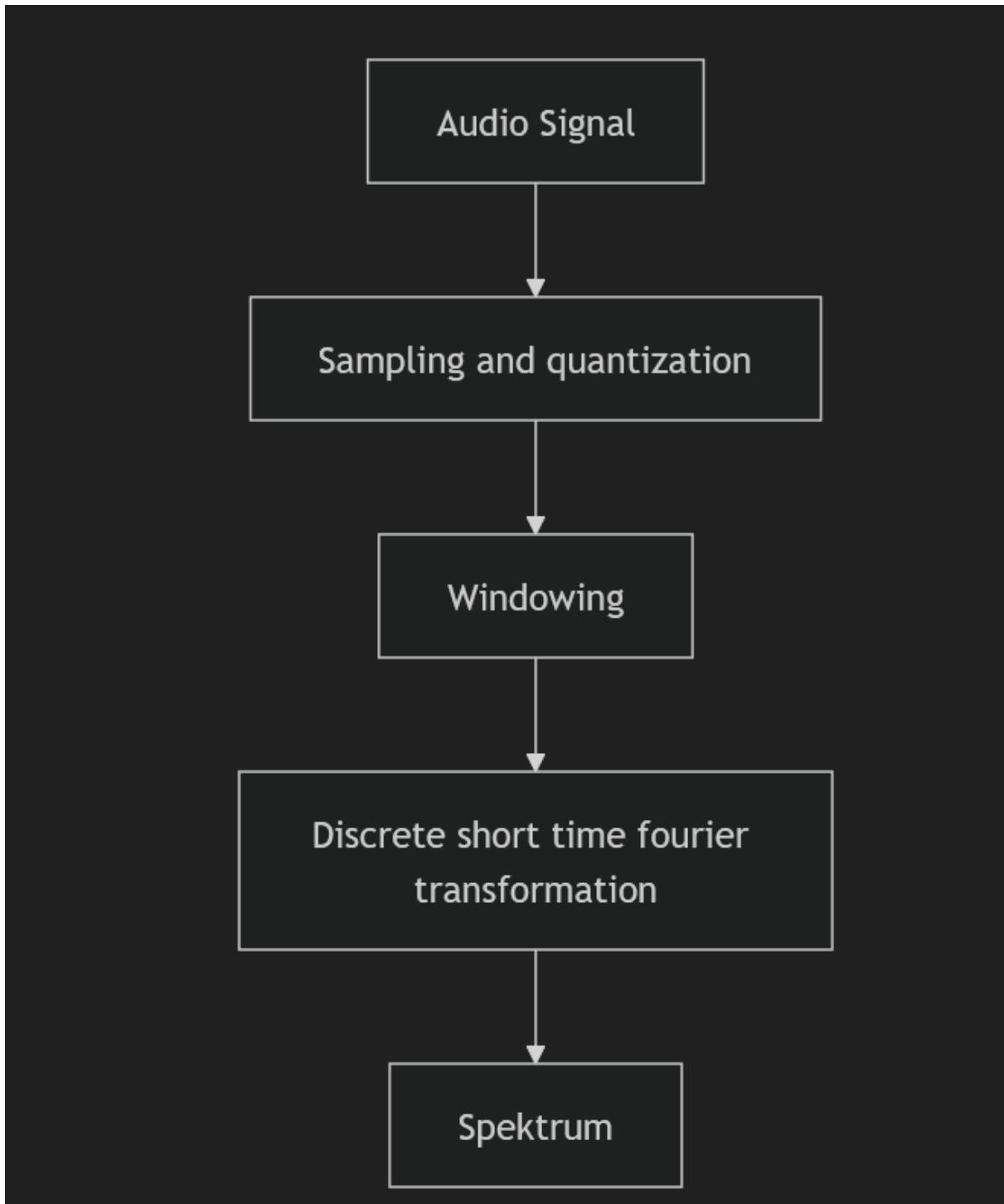


# Tutorium 6

Tags

## Aufgabe 1

```
graph TD
    a[Audio Signal] --> b[Sampling and quantization]
    b --> Windo
    wing --> d[Discrete short time fourier transformation]
    d --> Spektrum
```



Analog - Digital: Abtastrate und Frequenz beachten

DiscreteShortTimeFourierTransformation gibt ein Spektrum → Daraus Powerspektrum machen

## Aufgabe 2

**a**

Man erkennt Pausen, die Lautstärke aber NICHT die Wortgrenzen. Eventuell lässt sich aus ganz kleinen Wellenformen noch die Anwesenheit leichter Hintergrundgeräusche erahnen.

→ gut für voice activity detection, wird z.B. von Sprachassistenzen genutzt um zu erkennen ab wann er aktiviert bzw. deaktiviert werden muss

→ also: Amplitude über die Zeit, Segmentierung in Sprache/Stille möglich

**b**

Nein, einzelne Laute können nicht heraus analysiert werden

## Aufgabe 3

**A**

ist natürliche/spontane Sprache, also z.B. ein normales Gespräch ohne viele Pausen

**B**

ist isolierte Sprache, nach jedem Wort wird eine längere Pause gemacht, war nötig für das erst Spracherkennungssystem in 1995, das System hieß *Dragon Dictate*, musste isoliert eingesprochen werden und hat dann Stück für Stück das Eingesprochene transkribiert

→ isolierte Sprache war nötig um die Wörter von einander trennen zu können, da die Trennung mit voice activity detection (VAD) stattfinden konnte

## Aufgabe 4

Im Spektrogramm: Das bunte ist der Anteil der Frequenz in der Sprache (das Power Sprektrum)

Info: Frikativ ist die Frequenz höher, da durch das Pressen der Laute zwischen Zähnen und Lippen viele Wirbel entstehen

Im Gegensatz dazu: Vokale können sehr ungehindert rauskommen

- Erstes Segment (bis ca. 0.3 Zeit): Stille
- Zweites Segment (0.3 bis 0.55): Frikativ, da viele hohe Frequenzen
- Drittes Segment: Plosiv, also am Anfang einen glottalen Verschluss (es kommt also keine Luft durch), daher die Unterbrechung der Energie, sichtbar durch den schwarzen Streifen)
- Viertes Segment (ca. 0.65 nis 1): Sprache, Vokal, die Frequenz der Stimmenbänder gut sichtbar, hier ein U da der Rachenteil sich etwas verschließt und dadurch den Frequenzbereich runterdrückt
- Fünftes Segment: Nasal (ähnlich wie Vokal nur weniger Energie)
- Sechtes Segment: Stille

## Aufgabe 5

- Echo
  - Grund: Großer Raum
  - Lösung: Schalldämpfung, Teppiche etc.
- Hintergrundgeräusche (Gemurmel von Leuten)
  - Lösung: Tür zu
- Schalldruckwelle ist komplett ausgeschlagen, man sieht also nicht die peaks weil der Amplitudenbereich komplett ausgenutzt wurde (Übersteuert, Clipping), die reale Höhe der Peaks kann also nicht erkannt werden
  - Lösung: Lautstärke verringern oder Nahsprechmikrofon nutzen

- ganz unten kleiner gelber Streifen: gibt eine tiefe Frequenz die die ganze Zeit ausgeschlagen hat, entweder Sensor kaputt oder eventuell elektrisches Brummen
  - Lösung: gleichbleibende Frequenzen einfach rausfiltern, Nachteil: wichtige Frequenzen könnten verloren gehen
  - besser: Schauen, dass die Quelle des Grundrauschen beseitigt wird

## Übung 1-D-Frequenzraum (Audio)

- Grundklang am Anfang sichtbar (die Oberschwingungen der Glocke)
- Rauschen dass erst laut und dann etwas leiser/tiefer wird (am Peak erkennbar)
- Sprache, Phonanten der Vokale erkennbar
- tiefes Brummen erkennbar (gelbe Stellen)

## Teil 2

- Musik: das letzte
- Glas bersten beim Zerbrechen des Schiffs: das dritte
- Sprache und Musik abwechselnd (Anfang des Trailers): Das zweite (das gelbe ist der jauchzer)
- Der Start am Anfang (crescendo) mit Tush und Übergang zum Titanic-Motiv: erstes Bild



1D und 2D Fourier-Transformation Beispiel wird nächstes mal an der Tafel gemacht

