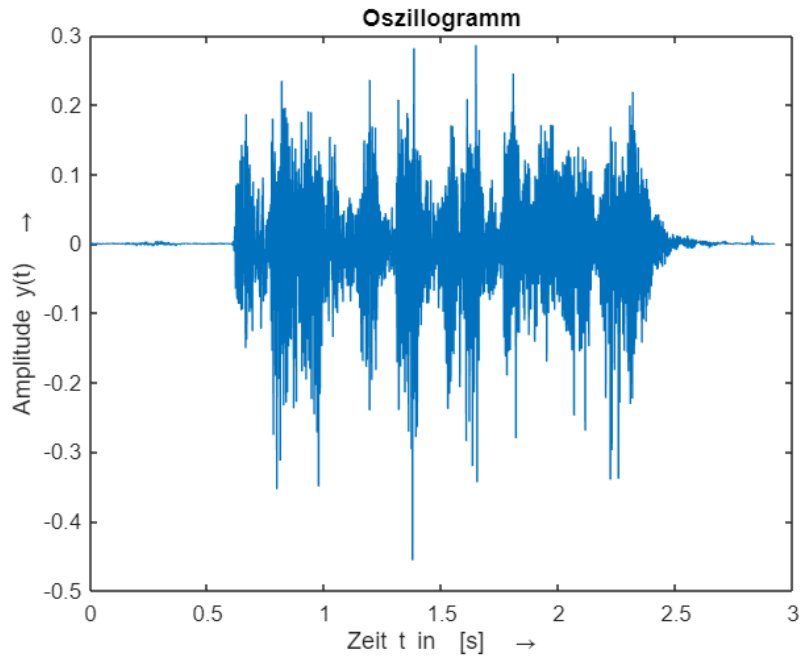


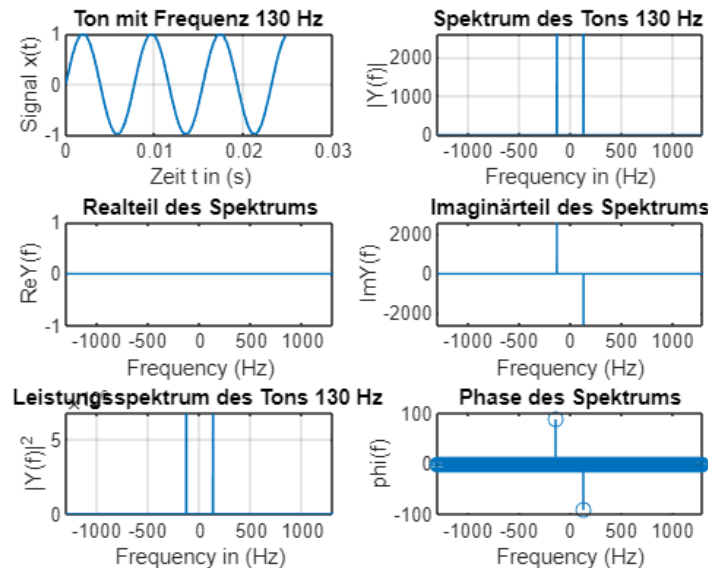
Labor 1: Sprachsignale plotten, laden und erzeugen



Optionale Tools: Audiotonic Pro, [PyCharm](#) als IDE
Python-Bibliotheken: numpy, scipy.io, matplotlib, ipython.display, sounddevice

- Um die Installation und die Grundlagen von Jupyter Notebooks zu lernen, können Sie das folgende Tutorial durchführen: <https://www.dataquest.io/blog/jupyter-notebook-tutorial/>
- Erzeugen Sie eine Audio-Datei auf ihrem PC mit folgendem Inhalt: "Dies ist eine Suchmaschine". Verwenden Sie als Abtastrate $f_a = 16\text{kHz}$. Erstellen Sie anschliessend ein Jupyter-Notebook das die erstellte Audio-Datei lädt, abspielt und das Sprachsignal als Funktion der Zeit plottet. Können Sie im geplotteten Sprachsignal Teile ihres Satzes wieder erkennen? Analysieren Sie ihr Sprachsignal schrittweise und beschreiben Sie ihr Ergebnis textuell.
- Generieren Sie jetzt ein harmonisches Signal bestehend aus dem Kammerton ($f_0=440\text{Hz}$) und seiner 2. und 3. Harmonischen. Die Amplitude A und die Zeitdauer t_d seien gegeben durch: $A= 1.0$ und $t_d=1\text{s}$. Tasten Sie das erstellte Signal mit einer Abtastfrequenz $f_a = 20 * f_0$ ab und speichern Sie es in einer wave-Audiodatei ab. Plotten Sie das resultierende Oszillogramm und den Kammerton. Lesen Sie die Audio-Datei ein und geben Sie den Klang aus.
- Verändern Sie jetzt die Phase der drei Einzeltöne mit einer Zufallsfunktion und plotten Sie den Klang als Zeitfunktion und geben das Audiosignal aus. Verändert sich der Klang sichtbar und hörbar?
- Mit welchem Schallsignal können sie den Kammerton komplett auslöschen. Generieren Sie in ihrem Programm das resultierende Signal und plotten und spielen Sie dieses ab.
- Erstellen Sie ein Programm zur Modellierung einer periodische Rechteckfunktion ($f_0 = 1\text{Hz}$, Breite=0.5s, $f_a = 8\text{kHz}$) durch Überlagerung aus seinen ersten 9 harmonischen Komponenten. Plotten Sie die so erzeugte Rechteckfunktion und zusätzlich die 1. Harmonische und die 9. Harmonische. Erklären und beschreiben Sie den Zusammenhang zwischen den 3 Funktionen.
- Wie klingt eine periodische Rechteckfunktion mit $f_0 = 440\text{Hz}$? Vergleichen Sie den Klang mit einem reinen sinus-Ton $f_0 = 440\text{Hz}$?

Labor 2 : Diskrete Fouriertransformation von Sprachsignalen

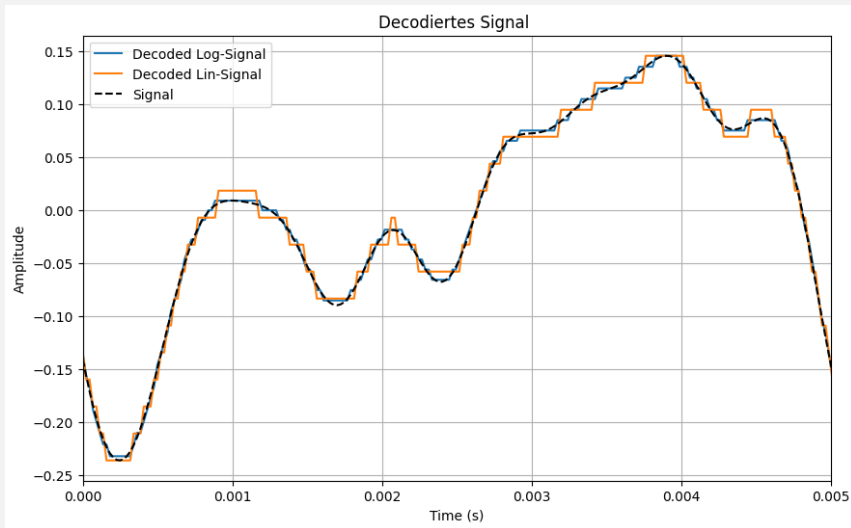


- a. Plotten und hören Sie sich den Grundton $y(t)$ ihrer Stimmbänder ($f_{0\text{male}} = 130\text{Hz}$, $f_{0\text{female}} = 200\text{Hz}$; $x_0 = 1.0$; $t = 2\text{s}$) an, und berechnen Sie anschließend die mittlere Leistung des Schallsignals in der Zeitdomäne.

Bestimmen Sie dann das Spektrum $Y(f)$ und plotten Sie den Betrag $|Y(f)|$ und das Betragsquadrat $|Y(f)|^2$. Erklären Sie die beiden Diagramme und berechnen Sie die mittlere Leistung des Schallsignals in der Frequenzdomäne. Stimmen die Leistungswerte im Zeit- und Frequenzbereich überein? Plotten Sie zusätzlich den Real- und Imaginärteil und das Phasen-spektrum. Erklären Sie auch diese Diagramme.

- b. Verkürzen Sie nun die Länge des Signals (1 Periode, $\frac{1}{2}$ Periode, $\frac{1}{4}$ Periode, ...) und beobachten und erklären Sie das Verhalten des Spektrums.
- c. Analysieren Sie die von Ihnen in Labor 1 erstellte Sprachbotschaft "Dies ist eine Suchmaschine". Laden Sie das Signal und plotten Sie den Signalverlauf, das zugehörige Leistungsspektrum in [db] und die Phase. Interpretieren Sie die Diagramme.
- d. Schneiden Sie anschließend mittels eines Audioprogrammes die beiden Wörter "Dies ist" und das Wort "Dies" aus ihrem Sprachsignal aus und speichern diese in einer separaten Audio-Datei. Suchen Sie sich in öffentlich verfügbaren Sprachkonserven Audiodateien für den Vokal "i" und den Konsonanten "s". Berechnen Sie anschließend die zugehörigen Leistungsspektren und vergleichen Sie diese mit dem Leistungsspektrum der gesamten Sprachnachricht. Erklären Sie ihre Beobachtungen. Plotten Sie den jeweiligen Signalverlauf inklusive dem quadratischen Mittel (rms: root mean square) der Amplitude des Signals. Was beobachten Sie?
- e. Berechnen Sie die Faltung zwischen einer Sinusfunktion ($f_0 = 1\text{Hz}$, $x_0 = 1$) und der Impulsantwort $e^{-t/2}$ im Bereich $t \in [0, 4 \cdot \pi \text{s}]$. Interpretieren Sie das Ergebnis.

Labor 3 : Quantisierung von Sprachsignalen

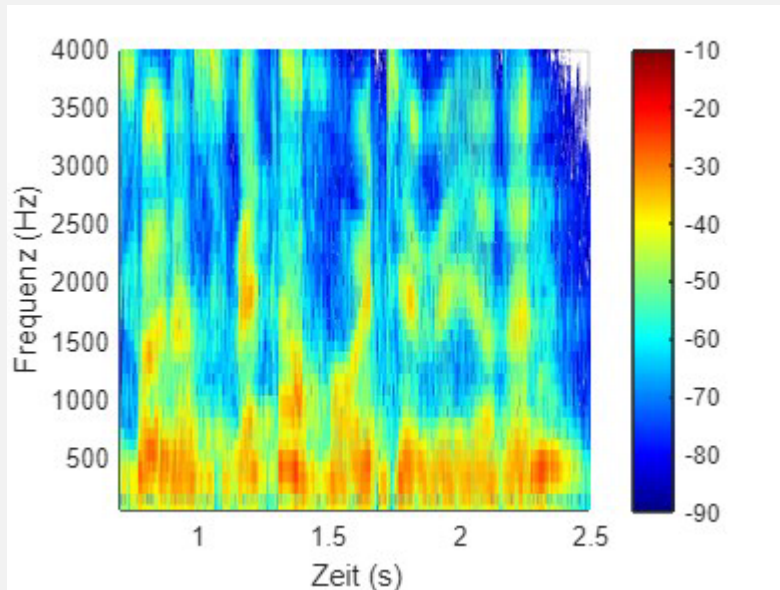


- Im folgenden verwenden Sie den Kammerton ($f_0 = 440 \text{ Hz}$, $x_0 = 1$) als Referenzsignal für die Quantisierung.
Plotten Sie die folgenden Funktionen x , $\sin(x)$, $\ln(\sin(x))$, $\ln(|\sin(x)|)$ und $\ln(|\sin(x)| + 1)$ und vergleichen deren Kurvenverlauf.
- Generieren Sie anschließend ein sogenanntes sweep-Sprachsignal mit $f_0 \in [200\text{Hz}, 1,6\text{kHz}]$ und $f_1 \in [500\text{Hz}, 4,0\text{kHz}]$ mit einer Dauer von $t_d = 3\text{s}$ und einer zufällig gewählten Phase φ .
Recherchieren Sie den Einsatz von sweep-Signalen.

$$x(t) = \sum_{i=1}^8 \left(x_0(i) \cdot \sin \left(2\pi \cdot \left(f_0(i) \cdot t + (f_0(i) - f_1(i)) \cdot \frac{t^2}{2 \cdot t_d} \right) + \varphi(i) \right) \right)$$

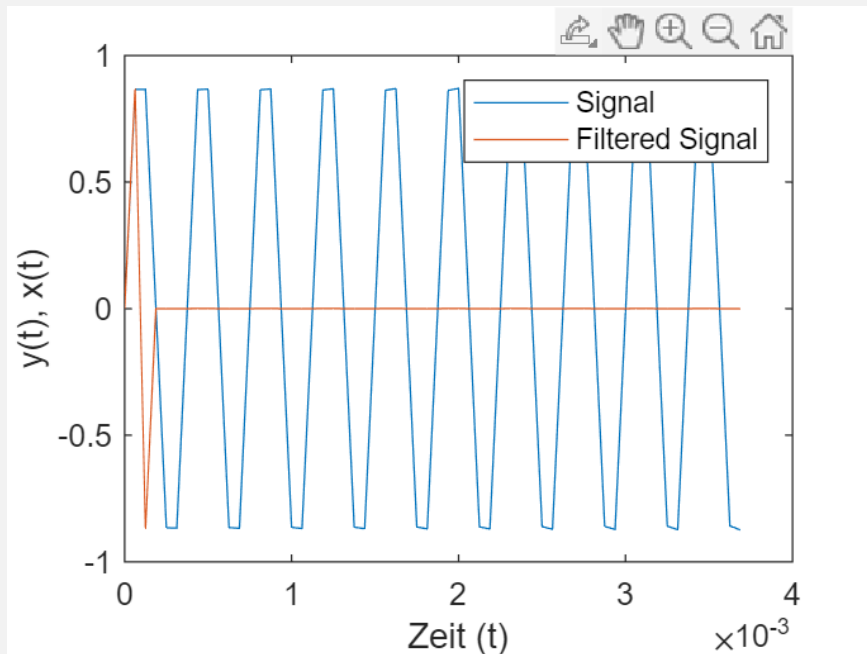
- Erstellen Sie jetzt einen logarithmischen und linearen Quantisierer und kodieren das sweep-Sprachsignal mit 2Bit und 8Bit. Plotten Sie jetzt die kodierten Signale beider Verfahren.
- Rekonstruieren Sie jetzt aus den kodierten Signalen das ursprüngliche Signal und plotten diese als Funktion der Zeit. Berechnen Sie anschließend das Quantisierungsrauschen und den SNR-Wert als Funktion der Zeit und stellen Sie diese geeignet grafisch dar.
- Erstellen Sie jetzt ein eigenes Sprachsignal z.B.: "ich liebe es zu reisen" mit einer Quantisierung von 32Bit und einer Abtastrate von 44,1kHz. Laden Sie das Signal und plotten Sie den Signalverlauf und das zugehörige Leistungsspektrum in [dB].
- Kodieren Sie jetzt das geladene Sprachsignal mit unterschiedlicher Bitanzahl $q = (2, 4, 8, 16)$ für die lineare und logarithmische Quantisierung. Plotten Sie die kodierten Signale und geben Sie die kodierte Sprachnachricht für die unterschiedlichen Verfahren auf dem Lautsprecher ihres PCs aus. Diskutieren Sie die beobachteten Ergebnisse.
- Berechnen Sie den SNR-Wert für jeden Zeitwert und bestimmen Sie den zugehörigen mittleren SNR-Wert. Erstellen Sie auch hier geeignet Diagramme und diskutieren Sie die Werte für den mittleren SNR.

Labor 4: Spektrogramm, AKF, Periodogramm, $H(z)$



- Berechnen und plotten Sie das Spektrogramm eines von Ihnen erstellten Audiotextes ("Dies ist eine Suchmaschine."). Erstellen Sie ein Schmalband- und Breitband-Spektrogramm. Wann erfolgt ein sichtbarer Übergang zwischen dem Schmalband- und dem Breitbandspektrum? Erklären Sie das Verhalten? Extrahieren Sie anschließend Formanten aus ihrem Spektrum und versuchen Sie diese einzelnen Vokalen zuzuordnen. Welchen Zeitbereichen können Sie Konsonanten zuordnen? Warum ist diese Zuordnung eindeutig?
- Extrahieren Sie aus ihrer Sprachnachricht jeweils einen Vokal und einen Konsonanten und berechnen die zugehörige Autokorrelationsfunktion. Erklären Sie die Unterschiede zwischen der AKF eines Konsonanten und der AKF eines Vokals.
- Bestimmen und plotten Sie das Periodogramm und das Welch-Leistungsspektrum eines Konsonanten. Vergleichen Sie die beiden Diagramme.
- Modellieren Sie nun weißes Gauß'sches Rauschen mit den Parametern ($\mu = 0, \sigma = 1$). Plotten Sie dazu ein Oszillogramm der Länge von 1s mit einer Abtastrate $f_a = 8000\text{Hz}$.
- Berechnen und plotten sie die AKF des soeben modellierten Rauschsignals. Berechnen Sie zusätzlich aus der AKF die spektrale Leistungsdichte. Vergleichen und diskutieren Sie die Diagramme.
- Bestimmen und plotten Sie jetzt das Welch-Leistungsspektrum des weißen Gauß'schen Rauschen und vergleichen Sie es mit der in vorheriger Teilaufgabe bestimmten Leistungsspektrum.

Labor 5: Filter und z-Transformation für Sprachsignale



- Bestimmen Sie analytisch die Systemfunktion $H(z)$ eines FIR-Filters mit den Nullstellen ($z=\{1, e^{\pm i\pi/3}\}$) und den Polstellen ($z=\{0,0,0\}$). Bestimmen Sie die Filterfrequenzen als Funktion der Abtastfrequenz f_a des zeitlichen Signals.
- Plotten Sie $|H(z)|$ über der komplexen Ebene ($\text{Re}(z)$ - $\text{Im}(z)$) und erstellen Sie einen Zero-Pole-Plot für $H(z)$. Vergleichen Sie die beiden Zeichnungen.
- Ergänzen Sie in beiden Schaubildern den Einheitskreis in der komplexen Ebene und diskutieren Sie den Einfluss der Pol- und Nullstellen.
- Plotten Sie die zeitliche Impulsantwortfunktion $h(n)$ als Funktion der Zeit für $f_a = 16$ kHz.
- Formulieren Sie anschließend die Systemantwort $y(n)$ in Form einer Differenzengleichung.
- Simulieren und plotten ($f_a = 16$ kHz) Sie die Systemantworten $y(n)$ für die Töne ($f_0 = 1000\text{Hz}, 2000\text{Hz}, 2666\text{Hz}$). Bei welcher Frequenz f_0 wird der eingehende Ton komplett gefiltert.
- Berechnen Sie den zugehörigen Frequenzgang $H(f)$ und plotten Sie dessen Betrag und die Phase. Vergleichen Sie ihr Ergebnis mit Teilaufgabe f.
- Berechnen Sie für einen FIR-Filter mit Eingangssignal $x(n)$ und der Impulsantwortfunktion $h(n)$

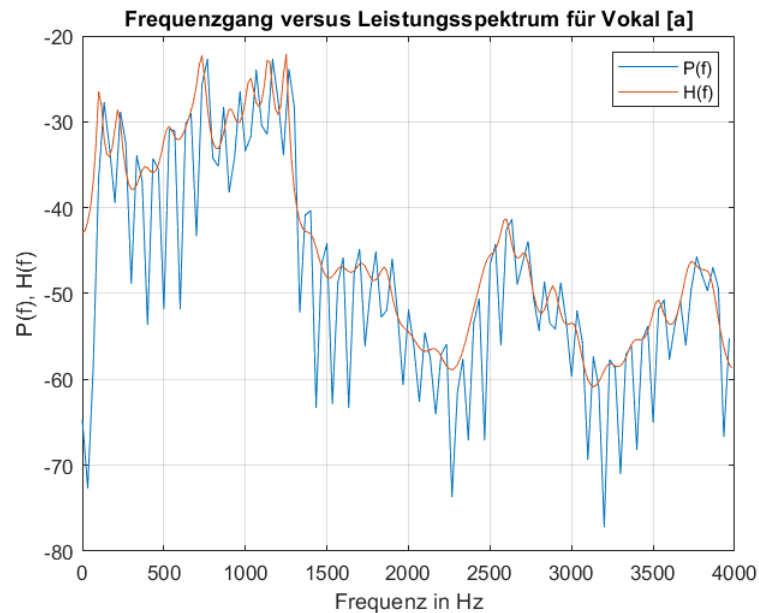
$$x(n) = \delta(n-1) - \delta(n-2) + \delta(n-3) - \delta(n-4)$$

$$h(n) = \delta(n) + 2 \cdot \delta(n-1) + 3 \cdot \delta(n-2) + 4 \cdot \delta(n-3)$$

jeweils deren Z-Transformierte $X(z)$, $H(z)$ und die z-Transformierte $Y(z)$ der Systemantwort.

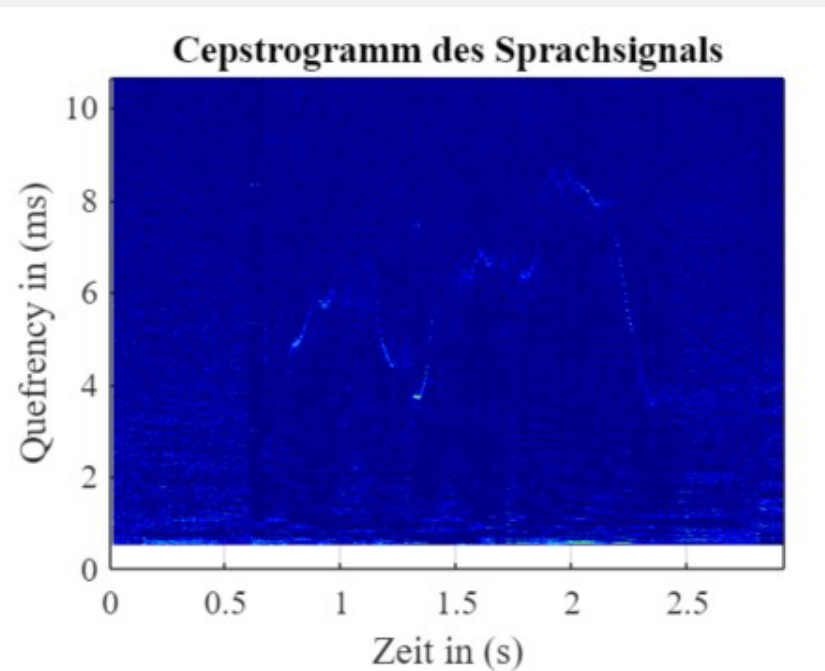
- Bestimmen Sie aus $Y(z)$ die Systemantwort $y(n)$ und plotten Sie alle 3 Signale ($x(n)$, $h(n)$, $y(n)$) als Funktion der Zeit für eine Abtastfrequenz $f_a = 16$ kHz.
- Bestimmen Sie die Null- und Polstellen von $H(z)$ und plotten Sie diese in einem zero-pole-Diagramm. Bestimmen Sie auch den Frequenzgang $H(f)$. Diskutieren Sie die beiden Diagramme.

Labor 6: Sprach-Analyse und Sprach-Synthese mit LPC



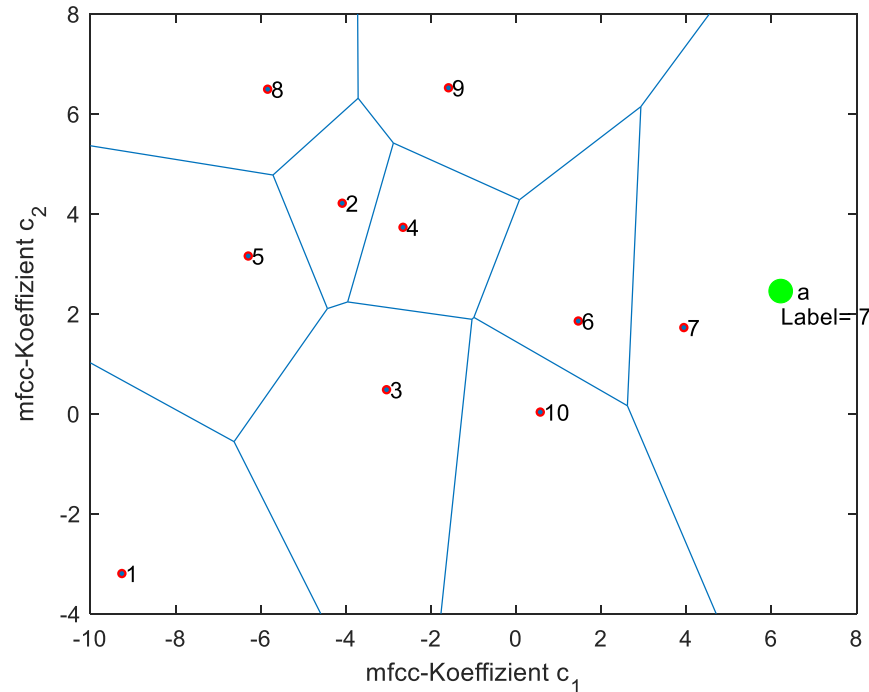
- Führen Sie eine LPC-Analyse für den Vokal [a] und den Konsonanten [f] durch. Schneiden Sie hierzu aus einem Sprachsignal eine ca. 20ms lange Sequenz aus, die den jeweiligen Laut enthält. Bestimmen Sie den inversen und den Synthesefilter. Zeichnen Sie die Zero-Pole-Darstellung und interpretieren Sie das Ergebnis.
- Berechnen Sie anschließend den Frequenzgang des Synthesefilters und vergleichen Sie das Ergebnis mit dem Leistungsspektrum ihres Signals. Variieren Sie die Prädiktor Ordnung M und beschreiben das beobachtete Verhalten.
- Führen Sie nun eine LPC-Analyse für einen ganzen Satz durch ("Dies ist eine Suchmaschine, die die neuesten Erkenntnisse liefert."). Zeichnen Sie das Signal mit einer Abtastrate von 8kHz auf. Bestimmen Sie anschließend mit einer Kurzzeit-LPC-Analyse für eine Fensterbreite ($w=200$, $w_{\text{shift}}=w/2$) den inversen Filter A ($M=12$) und speichern diesen pro Zeit Slot als Spalte in einer Matrix.
- Bestimmen Sie mittels der Autokorrelationsfunktion den zeitlichen Verlauf der Grundfrequenz F_0 . Plotten Sie F_0 und die Maxima der AKF als Histogramm im zeitlichen Verlauf.
- Führen Sie nun eine LPC-Synthese mittels den in Teilaufgabe c pro Zeitabschnitt bestimmten Filters A durch. Setzen Sie ihr Signal abschnittsweise unter Verwendung der Grundfrequenz F_0 und der Gesamtenergie E des Signals zusammen. Ermöglichen Sie dabei die Variation der Grundfrequenz F_0 sowie die Variation der zeitlichen Signallänge L . Plotten Sie das ursprüngliche und das generierte Signal und vergleichen Sie beide. Geben Sie ihr generiertes Signal auf einem Lautsprecher aus und diskutieren Sie die erreichte Sprachqualität mit dem Original. Verändern Sie die Parameter M , F_0 und L und diskutieren Sie die Ergebnisse.

Labor 7: Cepstrale Analyse und Cepstrogramme.



- Berechnen Sie für einen stimmhaften Laut (Vokal "a") und einen stimmlosen Laut (Konsonant "s") das geglättete Spektrum aus einer unterschiedlichen Anzahl an cepstral Koeffizienten. Verändern Sie die Anzahl der verwendeten Koeffizienten und vergleichen Sie deren Einfluß.
- Führen Sie nun eine LPC-Analyse für die näherungsweise gleiche Anzahl an Koeffizienten durch, und vergleichen das LPC-Spektrum mit dem aus dem Cepstrum gewonnen Spektrum für den stimmhaften und stimmlosen Laut. Berechnen Sie hierzu den Prädiktionsfehler $e(n)$ und die mittlere Amplitude des Fehlersignals. Welche Unterschiede stellen Sie zwischen den beiden Methoden fest? Begründen Sie diese?
- Plotten Sie den Prädiktionsfehler $e(n)$ zusammen mit dem Audiosignal für "a" und "s" jeweils in einem Oszillogramm. Wie gut ist ihre Prädiktion?
- Nehmen Sie mit einem Audioprogramm einen von Ihnen gesprochen Satz ("Wie wird das Wetter heute werden?") auf. Berechnen Sie anschließend für diesen Satz das zugehörige Cepstrogramm und visualisieren Sie den Verlauf der Grundfrequenz F_0 entlang des gesprochenen Satzes.
- Erstellen Sie sich nun eine Mel-Filterbank bestehend aus 24 Filtern und berechnen und plotten Sie das zugehörige Mel-Cepstrum für den Konsonanten "s". Bestimmen Sie aus dem Mel-Cepstrum das geglättete Mel-Spektrum des Konsonanten "s". Vergleichen Sie die so erhaltene Kurve mit der aus dem Cepstrum erhaltenen Kurvenverlauf.

Labor 8: Vektorquantisierung – Spracherkenner für Laute



1. Berechnen Sie auf Basis der Mel-Cepstren von 180 Audiodatei, die von 10 verschiedenen Lauten erzeugt wurden (a e f i l m n o s u) mittels dem K-Means-Algorithmus das zugehörige Codebook. Welchen Wert für K wählen sie? Speichern Sie das Codebook geeignet ab. Plotten Sie anschließend in der mfcc(1) – mfcc(2)-Ebene ihre Trainingsdaten und die ermittelten Codebook-Vektoren.
Hinweis: Die Mel-Cepstren können Sie der Datei **mfccs_db.xlsx** im Moodle-Kurs entnehmen. Zeile 1 -13 in der dort enthaltenen Matrix beschreiben die Mel-Koeffizienten (1-13). Zeile 14 enthält einen Codewert für jeden Laut: (a : 1, e : 2, f : 3, i : 4, l : 5, m : 6, n : 7, o : 8, s : 9, u : 10)
2. Bestimmen die relative Anzahl der Trainingsvektoren pro Voronoi-Bereich und erstellen Sie einen 3-dimensionalen Plot aus Codebook-Vektoren und berechneter Häufigkeit. Interpretieren Sie das Ergebnis.
3. Erstellen Sie jetzt eine Funktion die für einen Eingabe-Vektor (eingehendes Sprachsignal) den Code-Vektor mit der geringsten Distanz bestimmt und den Index des Codevektors (=Position in Codebook) als Wert zurückgibt.
4. Bestimmen Sie anschließend die Zuordnung Codebook-Nummer zu Einzellaut, indem Sie die Trainingsvektoren als Input verwenden und die bestimmte Code-Nummer ausgeben lassen.
5. Erstellen Sie sich jetzt Audiodaten für die einzelnen Laute (a e f i l m n o s u). Die Länge der Audiodaten sollte 30ms betragen und mit 8kHz gesampled werden. Erstellen Sie anschliessend eine Funktion die für eine eingehende Audiodatei die Mel-Cepstren bestimmt. Die so bestimmten Mel-Cepstren verwenden Sie um den Code-Vektor mit der niedrigsten Distanz zu bestimmen. Sie haben jetzt ein Spracherkennungssystem für Laute.