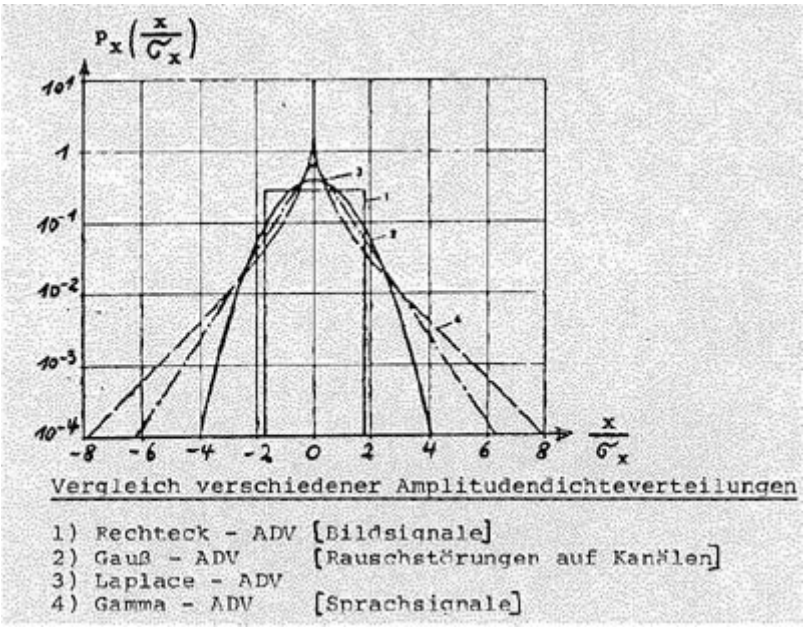
Sprachsynthese und –erkennung
Aufgabe 3

3.1 Statistische Beschreibung von Sprache



Schreibe ein Matlab-Skript zur Berechnung eines Histogramms für eine Sprachdatei (Wertebereich -1 -> 1*, Intervallgröße 0,1*). Zeichne für jedes Gruppenmitglied mit PRAAT einen Satz (ca. 10 Wörter) auf und berechne jeweils die Histogramme (Pausen vor und hinter dem Sprachsignal mit PRAAT abschneiden). Zeichne diese für die einzelnen Dateien graphisch in dasselbe Diagramm (x-Achse: Amplituden-Wertebereiche, y-Achse: Prozentangaben). Wie unterscheiden sich die ADVs der einzelnen Dateien? Aus dem Histogramm läßt sich auch ablesen, wie gut die Aufnahmen ausgesteuert sind und ob das Signal einen DC-Offset hat. Wie?

3.2 Phoneme

Phoneme sind die kleinsten bedeutungsunterscheidenden Elemente einer Sprache. In dieser Übung werden wir sogenannte Minimalpaare, d.h. Wörter, die sich genau in einem Phonem unterscheiden, suchen, diese aufzeichnen und die Oszillogramme, d.h. die Zeitsignale, segmentieren und vergleichen. Genaugenommen sind es nicht immer Minimalpaare, sondern oft Tripel und Quadrupel etc. die sich jeweils an einer Stelle unterscheiden (vgl. Anne vs. Amme vs. Affe vs. Asche etc.). Wer findet die meisten Varianten? ALLE BEISPIELE VON JEDEM GRUPPENMITGLIED SPRECHEN LASSEN!!!
Die "Minimalpaare" sollen folgendes Aussehen haben ("C" steht hier für Konsonant, "V" für Vokal):

(a) einsilbig mit wechselndem Vokal: CVC (insgesamt fünf verschiedene)
(b) zweisilbig mit wechselndem Konsonanten: VCV (insgesamt fünf verschiedene)

Am besten platziert man alle Beispiele zu einem Bildungsmuster in einer Datei, das erleichtert den Vergleich.

"Segmentieren", d.h. die Lautgrenzen markieren, werden wir wieder mit dem Labeltool in PRAAT (Praat objects->annotate ->To TextGrid). Das Grid dann mit dem Wave-File zusammen für spätere Anwendungen abspeichern (mit dem Cursor Wave und TextGrid invertieren, unter "Write->Write to binary file" als "praat.Collection" abspeichern, eindeutigen Namen vergeben!).

Wenn sämtliche Beispiele markiert sind, wollen wir die Unterschiede zwischen den Vokalen, bzw. Konsonanten im Oszillogramm und im Spektrogramm untersuchen. In welchen Merkmalen unterscheiden sich die von euch gewählten Konsonanten und wie drücken sich diese im Oszillogramm und Spektrogramm aus? In welchem Frequenzbereich befindet sich z.B. die meiste Energie?

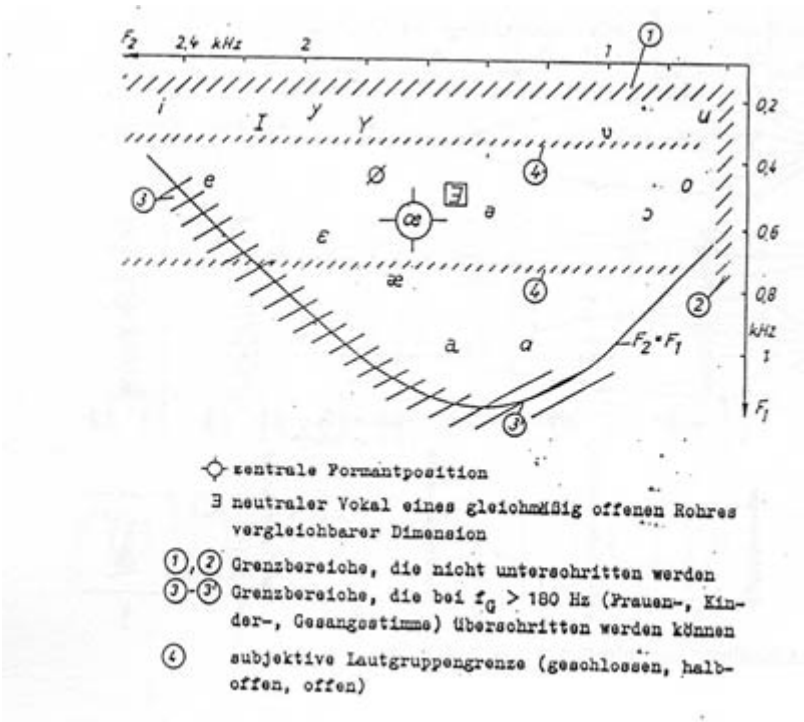
Beachtet besonders genau die Übergänge V->C und C->V. Welche Artikulationsart, bzw. -stelle bezeichnet die von euch gewählten Konsonanten? Am besten probiert man das an sich selbst aus, d.h. die Lautfolge langsam sprechen und darauf achten, wie sich die Zunge im Mundraum bewegt (vor->zurück, hoch->tief) etc.

Wie unterscheiden sich die von euch gewählten Vokale (Zungenposition, Länge -> ausmessen, Lippenrundung +/-)? Vergleicht die Oszillogramme/Spektrogramme verschiedener Sprecher, ähneln sich diese für den gleichen Laut?

Protokoll: Beispiele als Text und Transkription, Plot der Beispiele (mit TextGrid) und Wave-Audio, kurze Beschreibung zu den Eigenschaften der variierten Laute. Plots der Lautübergänge mit den gemachten Beobachtungen.

3.3 Formanten

Die Bereiche hoher spektraler Energie eines Lauts bezeichnet man auch als "Formanten". Speziell die Vokale lassen sich relativ gut durch ihren 1. und 2. Formanten, kurz F1 und F2 genannt, beschreiben. Das sind die Mittenfrequenzen des ersten und zweiten Energieschwerpunkts, die man in PRAAT als Kurvenzüge über der Zeit darstellen lassen kann (View->Formants). Diese resultieren aus den Resonanzfrequenzen des Vokaltrakts, die wiederum durch dessen Geometrie bedingt sind (Details in der Vorlesung!) und bei denen das glottale Anregungssignal gut verstärkt wird. Da es sich bei einem Formanten genau genommen um ein ganzes Frequenzband handelt, wird er durch seine Mittenfrequenz und seine Bandbreite beschrieben, in PRAAT als "formant" und "width" bezeichnet.
Wir erinnern uns an das Vokalviereck, bei dem links vordere Vokale und rechts hintere Vokale dargestellt werden. Die Lage der Vokale entspricht in etwa der Zungenstellung bei ihrer Artikulation. Wenn man in einem rechtwinkligen Koordinatensystem F1 auf der y-Achse und F2 auf der x-Achse abträgt, wobei abweichend von der üblichen Konvention der Nullpunkt rechts oben liegt, erhält man eine Art Formantkarte der Vokale, die in etwa dem Vokalviereck entspricht wie in der folgenden Abbildung.



Bei der Messung der Vokalformanten nimmt man sinnvollerweise als Ort der Messung die Lautmitte. Die folgende Tabelle gibt zur Orientierung die Formantfrequenzen einiger deutscher Vokale wieder. Hierbei handelt es sich um Mittelwerte, die über eine große Anzahl von Sprechern gemessen wurden.

Laut	F1 [Hz]	F2 [Hz]
i:	270	2290
I	390	1990
E	530	1840
@	660	1720
a	730	1090
U	300	1020
u:	300	870

Die Aufgabe in unserer Übung besteht nun daran, für die einzelnen Teilnehmer Formantkarten zu bestimmen. Dazu komplettieren wir die Vokale, die in den Beispielen aus Übung 3.2 noch fehlten, messen die Formantfrequenzen und tragen sie in ein Koordinatensystem ein (Plot der Formantfrequenzen für die Gruppenteilnehmer ins Protokoll). Dabei sollten die Messwerte aus PRAAT kritisch betrachtet werden, da es besonders bei hinteren Vokalen zu Fehlern, d.h. zur Vertauschung von Formanten kommen kann. Die Werte in der Tabelle dienen daher auch als Orientierung. Bei Diphthongen wie [aI] und [OY] hat die Messung in der Mitte keinen Sinn, sondern es müssen zwei Punkte, einer am Anfang und einer am Ende des Vokals gemessen werden.