Sprachsynthese und –erkennung Übung 2

Sprachsignale: Erste Schritte

Mit dem Headset und PRAAT soll von jedem Gruppenmitglied ein anderes dreisilbiges Wort aufgezeichnet werden. Das Wavesignal kann dann im Edit-Fenster angezeigt werden. Nun soll zunächst die phonetische Transkription nach der folgenden SAMPA-Notation erfolgen:

SAMPA	EXAMPLES (orthographic)
_	(silence, no signal)
p	Pier, aB
b	Bier, KraBBe
t	Tier, GraD
d	Dir, eDel
k	Kasse, taG
g	Gasse, eGal
f	Vogel, schlaF
v	Wasser, eVentuell
s	aSt, haSS
z	Sieb, beSen
S	Spät, aSCHe
Z	Genie, DSCHungel
x	naCH, doCH
C	diCH, honiG
h	Hut, aHorn
pf	PFerd, toPF
ts	Zwei, plaTZ
tS	kuTSCHe, Cello
m	Mut, haMMer
n	Nase, kaNNe
N	eNG, baNGe
l	Liebe, haLLe
R	Riese, kRaut
6	opER, deR
j	Jetzt, Jagd
aI	EIns, kAIser
OY	ÄUßerung, nEU
aU	AUf, schAU
@	sehEn, bEsagt
i:	Igel, bIEten
I	In, bItten
y:	Übung, hÜten
Y	Ypsilon, hÜtten
e:	bEten, schnEE
E	bEtten, gÄste
E:	Äsen, geblÄse
2:	Öfen, mÖgen
9	Öffnen, kÖnnen
u:	bUHlen, gUt
U	lUstig, bUtter
o:	Ofen, kOHl
O	Offen, tOpf
a:	wAr, wAHr
a	An, kAnn

Only rudimentary implemented:

E^ (= E~:)	bulletIN
a^ (= a~:)	pendANT
9^ (= 9~:)	parfUM
o^ (= o~:)	feuilletON

```
E~              IMpair
a~              pENdant
o~              nONchalant
```

Das heißt zum Beispiel, daß das Wort "Sprachlabor" wie folgt transkribiert wird: [SpRa:xlabo:6] (eckige Klammern sollen die Transkription kennzeichnen). Nun soll im Editor durch Anwahl mit der Maus die Länge der einzelnen Laute im aufgezeichneten Wort bestimmt werden. Dazu wählt man mit linkem Mausklick den Beginn des gewünschten Bereichs an und zieht mit der Maus nach rechts, so daß sich das rosaumrandete Fenster in der gewünschten Breite öffnet. Über dem rosa Rahmen erscheinen links und rechts Anfangs- und Endzeitpunkt des gewählten Bereichs, in der Mitte die Länge in Sekunden. Durch Abspielen (Button über Rahmen drücken) kann der Bereich nun wiedergegeben werden.

Darauf achten, dass in gesprochener Sprache oft Verschleifungen und Reduktionen auftreten! Laute, die nicht gesprochen wurden, können auch nicht ausgemessen werden.

Wir legen eine Tabelle in der folgenden Form an:

```
SAMPA           Anfang  [s]     Ende [s]        Dauer [s]       Grundfrequenz
(F0) [Hz]
S               1.5             1.7             0.2             -
p               1.7             1.8             0.1             -
R               1.8             2.0             0.2             150
...
...
```

Bei der Grundfrequenz handelt es sich um die Frequenz der Stimmbandschwingungen, die natürlich nur bei stimmhaften Lauten (stimmhafte Konsonanten und Vokale) existiert. Um sie zu bestimmen, geht man in die Mitte eines stimmhaften Lautes und zoomt so weit hinein, daß die Periodizität gut zu erkennen ist. Zur Orientierung kann man z.B. auffällige Amplitudenminima im Sprachsignal wählen und den Abstand zwischen zwei solchen charakteristischen Minima messen. Die Grundfrequenz ergibt sich dann aus dem Kehrwert (Beispiel: gemessener Abstand 4 ms -> Grundfrequenz: 1/(0.004 s)= 250 Hz).

Wenn uns diese Angaben vorliegen, können wir sie verwenden, um dasselbe Wort mit MBROLA synthetisieren zu lassen. Die Notation des von MBROLA verwendeten .pho-Files sieht folgendermaßen aus:

```
SAMPA   Länge in ms     % in Laut       Grundfrequenz   % in Laut
        Grundfrequenz
...
```

Man kann im Prinzip fast beliebig viele Stützstellen für die Grundfrequenz angeben, wir beschränken uns aber zunächst auf einen in der Mitte des Lautes. Für stimmlose Laute erübrigt sich die Angabe des Grundfrequenzwertes verständlicherweise. Damit ergibt sich für das obige Beispiele die folgende pho-Notation:

```
S       200
p       100
R       200     50      150
...
...
```

Gib das synthetische Wort wieder und vergleiche mit dem Original. Dazu kann das synthetische in ein WAV-File exportiert werden. Sieh dir das synthetische File in PRAAT an und kontrolliere ob die Lautlängen und die Grundfrequenz korrekt realisiert wurden. Wie unterscheidet sich die Synthese-Wellenform vom Original? Wie klingt die Synthese im Vergleich zum Original? Wie gut werden die Dauern und die Grundfrequenzkontur des Originals nachgebildet?

Es scheint einen Bug zu geben, bei dem MBROLA den letzten Laut manchmal "verschluckt". Abhilfe schafft die Einfügung einer Pause nach dem Wort (_ 100).

Verbessere das Ergebnis, in dem du je Laut drei Grundfrequenz-Werte mißt (am Anfang, in der Mitte und am Ende). Als nächstes sprich einen kurzen Satz und verfahre genau wie beim Einzelwort. Hier reicht ein Grundfrequenzwert pro Laut. Vergleiche die Sprachqualität von Einzelwort und Satz

Das Protokoll zur 2. Übung besteht aus folgenden Angaben:

Das gewählte Wort bzw. der kurze Satz, die dazugehörigen Wave-Dateien (tauchen Laute im Sprachfile gar nicht auf, die die Transkription erwarten ließ ?) und die Transkriptionstabellen, die entsprechenden pho-Files und die Synthese-Wave-Dateien. Bilder aus PRAAT (Wellenform, Spektrum , Pitch-Kontur, Textgrid), Angaben zum Vergleich natürlich/synthetisch. Wellenformen aller Dateien als Graphiken.