

Sprachsynthese und -erkennung - Übung 5

Prosodie

5.1) Diese Übung beschäftigt sich mit dem Thema Prosodie. Im Grunde ist es ein kleines Forschungsprojekt. Unter Prosodie verstehen wir die Sprachsignaleigenschaften, die eine Folge von Lauten betreffen wie *Intonation* und *Rhythmus*. Wir kennen bereits die Grundfrequenzkontur $f_0(t)$, mit der die Intonation am stärksten verknüpft ist. Sprachrhythmus ist gekennzeichnet durch die relative Dauer bzw. Stärke von Silben. Jedes mehrsilbige Wort hat einen solchen Rhythmus aus starken und schwachen Silben. Zum Beispiel kann der Rhythmus des Worts „ver-'steh-en“ als „kurz-lang-kurz“ oder „- - .“ beschrieben werden. „'“ kennzeichnet die Wortakzentsilbe.

5.2) Dieselbe Folge von Wörtern kann unterschiedliche Bedeutungen haben, abhängig von der Art, wie wir sie betonen. Zuerst suchen wir uns einen kurzen Satz mit der Struktur Subjekt-Prädikat-Objekt, wie zum Beispiel „Der Mann liest die Zeitung“. Diesen Satz können wir nun unterschiedlich betonen, je nach dem, nach welchem Satzteil gefragt wird, oder auch, worauf der sogenannte *Fokus* liegt:

- 1) Als Antwort auf die Frage: „Wer liest die Zeitung?“ - „*Der Mann* liest die Zeitung.“
- 2) Als Antwort auf die Frage: „Was tut der Mann mit der Zeitung?“ - „Der Mann *liest* die Zeitung.“
- 3) Als Antwort auf die Frage: „Was liest der Mann?“ - „Der Mann liest *die Zeitung*.“
- 4) Als Antwort auf die Frage: „Was passiert da?“ - „*Der Mann liest die Zeitung*.“

Beim letzten Beispiel 4) wird nach dem ganzen Satz gefragt, man spricht von einem „breiten Fokus“, bei den anderen (1-3), bei denen nur nach einem Teil gefragt wird, von einem „engen Fokus“.

Wir zeichnen jetzt die vier Varianten auf, indem ein Teilnehmer die Frage stellt und ein anderer antwortet. Achtet darauf, dass der Antwortende wirklich Aussagen spricht und nicht mit der Stimme am Ende hoch geht. Achtet auch auf gleichmäßigen Abstand des Mikrofons vom Mund und beim Sprechen auf eine einigermaßen gleichmäßige Lautstärke. Speichert nun die Aufnahmen als WAV-Dateien ab.

5.3) Nun segmentieren wir die Sätze im Praat TextGrid, aber nicht auf der Phonebene, sondern auf der Silbenebene. Das wäre in unserem Fall „Der-Mann-liest-die-Zei-tung“.

Nun messen wir die Silbendauern für alle Fälle, legen eine Liste an und vergleichen die Werte. Was hat der veränderte Fokus für eine Wirkung?

Dasselbe tun wir nun für die Grundfrequenz $f_0(t)$, indem wir den pitch-Wert für jede Silbe in Praat in der Mitte der Silbe messen und auch diesen Einfluss beschreiben.

Zum Schluss lassen wir uns die Intensität („Intensity“) anzeigen und notieren die dB-Werte in unserer Tabelle:

Fall 1)	Silbendauer [ms]	F0 [Hz]	Intensität [dB]
Der			
Mann			
liest			
die			
Zei			
tung			

5.4) Nun manipulieren wir die Grundfrequenz $f_0(t)$ in Praat, indem wir mit unseren Sounds *Manipulate->To Manipulation* aufrufen. Diese Manipulation beruht auf dem PSOLA-Verfahren, das wir bereits

kennengelernt haben. Im ManipulationEditor sieht man oben die markierten Perioden im Sprachsignal, darunter einen Bereich, in dem Pitch (f_0) angezeigt wird, und darunter einen Bereich zur Manipulation der Dauern.

Im Manipulation-Editor können wir Pitch auf einen konstanten Wert setzen und so die ursprüngliche Grundfrequenz $f_0(t)$ auf einen konstanten Wert setzen, indem wir alle PitchPoints löschen und dafür einen einzelnen PitchPoint setzen. Damit erhalten wir monotone Versionen unserer Sätze. Untersuche durch Abhören, ob die ursprünglichen Bedeutungen, d.h. der Fokus, erhalten geblieben sind. Speichere die monotonen Versionen als WAV-Dateien ab (File->Publish Resynthesis überträgt das aktuelle Audio in das ObjectWindow).

5.5) Jetzt nehmen wir Fall 4) und verringern die Geschwindigkeit, indem wir einen DurationPoint mit dem Wert 1,2 einsetzen. Wir erhöhen den Wert in Schritten von 0,2, bis wir bemerken, dass sich die Klangqualität deutlich verschlechtert. Woran liegt das? Speichere diesen Fall als WAV-Datei ab.

5.6) Zum Schluss nimm den Satz als Frage auf: „Der Mann liest die Zeitung?“ Vergleiche den Grundfrequenzverlauf mit der Aussage 4) und beschreibe die Unterschiede. Versuche nun, ob sich durch Übertragen des Pitch-Verlaufs die monotone Aussage in eine Frage ändern lässt. Das erfolgt ebenfalls im ManipulationEditor. Wir laden wieder die Aussage 4) und setzen in jede Silbe einen PitchPoint. Diese PitchPoints verschieben wir, bis die Aussage wie eine Frage klingt. Speichere auch diese Version als WAV-Datei ab.

Ins Protokoll: WAVs, sowie Oszillogramme und Spektrogramme aller Aufnahmen mit TextGrids für die vier Varianten der Aussage und die Frage, Tabellen mit den prosodischen Parametern – und natürlich Antworten zu den Fragen.