Das Annotieren von Intonation mit Wavesurfer

Kati Schweitzer

17. August 2006

1 Handhabung von Wavesurfer

1.1 Download

Wavesurfer ist an der KTH in Stockholm frei verfügbar. Man kann sich die Software (zB für daheim) unter
`http://www.speech.kth.se/wavesurfer/download.html`
herunterladen. Auf der Wavesurfer-Homepage gibt es auch ein Forum, das sehr gut betreut wird und mit dessen Hilfe sich viele Fragen beantworten lassen (`http://www.speech.kth.se/prod/wavesurferforum/phpBB2/index.php` bzw. durchklicken)

1.2 Erste Schritte

1.2.1 Starten

Wavesurfer wird (unter Linux) mit `wavesurfer&` bzw. `wavesurfer FILE &` auf der Shell gestartet. Es erscheint das Wavesurfer-Fenster. Hat man den Dateinamen noch nicht als Argument angegeben, kann man die Datei über das Menü starten (`File - Open...`).

1.2.2 Abspielen

Um das komplette File abzuspielen kann man entweder in der Menüleiste das `play` - Symbol auswählen, oder den Cursor an den Anfang des Fensters setzen und das Abspielen mit der Leertaste starten. Außerdem kann man mit der Maus eine Sektion des Signals auswählen und diese mittels des `play`-Symbols oder Drücken der Leertaste abspielen.

1.2.3 Zoomen

Es gibt in der Menüleiste vier Symbole in der Menüleiste fürs Zoomen. Mit ihnen kann man einzoomen, auszoomen, komplett auszoomem oder in eine vorher mit der der Maus markierter Selektion zoomen.

1.2.4 Konfigurationen

Das Grundprinzip beim Betrachten von Soundfiles mit Wavesurfer ist, dass man sich verschiedene Ebenen (*panes*) anzeigen lassen kann, je nachdem, welche Repräsentation des Sprachsignals benötigt wird. Um nicht jedes Mal die nötigen

Ebenen laden zu müssen, gibt es die Möglichkeit, vordefinierte oder eigene *Konfigurationen* zu wählen, die automatisch bestimmte Ebenen und Einstellungen laden. Zugang zu diesen Konfigurationen bekommt man über `RECHTSKLICK - apply configuration`. Man kann dann im erscheinenden Menü zwischen den verfügbaren Konfigurationen auswählen. Möchte man sich eine eigene Konfiguration erstellen, wählt man die gewünschten Ebenen und Einstellungen aus und speichert die Konfiguration über `RECHTSKLICK - save configuration`. Von da an ist sie ebenfalls in dem durch `RECHTSKLICK - apply configuration` aufzurufenden Menü zu finden.

1.2.5 Konfiguration fürs intonatorische Labeln

Beim Intonatorsichen Labeln ist die wichtigeste visuelle Unterstützung die Repräsentation der Pitch-Kontur, also der Verlauf der Grundfrequenz (F_0). Diese erhält man über

- `RECHTSKLICK - create pane... - Pitch Contour`

Eine sinnvolle Einstellung, die man vornehmen sollte ist hier das Sperren der Spur, da Wavesurfer standardmäßig so eingestellt ist, dass man bei mit der linken Maustaste die Kurve verändern kann. Dies führt oft zu ungewollten Veränderungen, da man mit der linken Maustaste ja auch einen Bereich zum Abspielen markieren kann - auch auf der Pitch-Kontur. Zum Sperren der Spur geht man folgendermaßen vor:

- `RECHTSKLICK - Properties` auf der Pitch-Kontur-Ebene
- Auswahl der Registerkarte `Data Plot`
- Aktivierten des Punktes `Lock data plot`

Wenn man sich dann schon in dem Menü befindet, kann man gleich noch ein bisschen Kosmetik betreiben und die Farbe der Kurve zB auf rot stellen:

- in der Registerkarte `Data Plot` unter `Plot column` direkt eine Farbe eintragen oder über `Choose...` eine auswählen

Neben der F_0-Kontur macht es Sinn, sich das Oszillogramm (*die Waveform*) anzeigen zu lassen:

- `RECHTSKLICK - create pane - Waveform`

Grundsätzlich wird fürs Labeln neben den Repräsentation des Sprachsignals noch eine Ebene für die Tranksriptionen benötigt. Das Labelfile-Format, das am IMS hauptächlich verwendet wird, ist das `Waves`-Format. Der Grund hierfür ist, dass für die Repräsentation der Wörter, Silben und Phone automatisch Labelfiles erzeugt werden können; da diese im `Waves`-Format sind, bietet es sich an, auch für die anderen Labelfiles diese Format zu wählen. Das Labelfile format, das von Wavesurfer standardmäßig verwendet wird, ist hingegen das `Wavesurfer`-Format. Das heißt, wann immer man eine neue Ebene für Transkriptionen erzeugt, muss erst das Labelfileformat umgestellt werden. Dafür geht man folgendermaßen vor:

- `RECHTSKLICK - create Pane`

- Auswahl `Transcription` (also eine Labelspur)
- `RECHTSKLICK - properties`
- Auswahl des Labelfile Formats `Waves` im obersten Punkt der Registerkarte `Trans 1`
- Schließen des Properties-Fensters mit `OK`

Auf der jetzt entstandenen Ebene kann man Transkriptionen in Form von Labels eintragen. Der einfachste (aber - zumindest für Labeln von Intonation - auch ineffizienteste) Weg, dies zu tun, ist es, die Spur anzuklicken und über die Tastatur ein Label einzutragen. Wesentlich sinnvoller ist es hingegen für Transkriptionen, bei denen man ein bestimmtes Inventar von möglichen Labels hat (wie z.B., im Fall von Intonations-Labelling, Akzenttypen oder Phrasengrenzen), diese möglichen Labels in den Einstellungen der Labelspur einzutragen. Danach können diese nämlich mittels `RECHTSKLICK` ausgewählt werden - man spart sich also eine Menge Tipparbeit. Zum Eintragen der Labels geht man folgendermaßen vor:

- `RECHTSKLICK - properties` (auf der Labelspur)
- Auswahl der Registerkarte `Trans 2`
- Eintragen der Labels in der Tabelle ganz unten (dafür evtl Anpassen der Spalten/Zeilenanzahl)
- Schließen des Fensters mit `OK`

Jetzt können auf der Spur, für die die Einstellung gemacht wurde, die entsprechenden Labels mittels Rechtsklick an der Stelle, an der das Label sitzen soll, ausgewählt werden.

Möchte man ein schon bestehendes Labelfile betrachten, so gibt es verschiedene Möglichkeiten. Will man das File nur einmalig anschauen, ist es das einfachste, es mittels

- `RECHTSKLICK - Load Transcription` auf einer Labelspur
- Auswahl des entsprechenden Files

in die entsprechende Spur hineinzuladen.

Will man hingegen häufiger das selbe Labelfile betrachten, bzw für verschiedene Soundfiles immer die gleichen Labelfiles betrachten, so bietet es sich an, dies in der Konfiguration einzutragen, so dass die entsprechenden Files immer sofort geöffnet werden. Hierfür hat man die Möglichkeit, für jede Labelspur anzugeben, welches Labelfile in ihr angezeigt werden soll. Dies funktioniert dann, wenn die entsprechenden Labefiles sich nur durch die Extension unterscheiden, d.h., wenn alle Files den selben Basename wie das Soundfile haben. Fürs Labeln von Intonation benötigt man konkret:

- ein Soundfile (Extension oder `.wav`), also `BASENAME.wav`
- ein File, das die Repräsentation der im Soundfile gesprochenen Wörter enthält: `BASENAME.words`

- ein File, das die Repräsentation der im Soundfile gesprochenen Silben enthält: `BASENAME.syl`
- ein File, das die Repräsentation der im Soundfile gesprochenen Laute enthält: `BASENAME.phones`.

Diese werden automatisch erstellt, sie dienen nur zur Unterstützung beim Lablen. Außerdem braucht man 2 Labelspuren für die zu erstellende Repräsentation der Intonation:

- ein File `BASENAME.accents`, in das die Labels für die Akzente eingetragen werden
- ein File `BASENAME.tones`, in das die Labels für die Grenztöne und Phrasenakzente eingetragen werden

Da diese Files immer geladen werden müssen, wenn man Labeln will, speichert man sie am besten in der Konfiguration ab. Hierfür geht man folgendermaßen vor:

- `RECHTSKLICK - Properties` auf der entprechenden Labelspur
- Auswahl der Registerkarte `Trans 1`
- Eintragen der gewünschten Extension unter `Label filename extension` (also zB `.words`)
- Schließen des Fensters mit `OK`

Hat man all diese Einstellungen vorgenommen, sollte man sich die Fenster in die gewünschte Größe ziehen und dann die Konfiguration abspeichern. Wählt man beim nächsten Mal diese Konfiguration aus, sind alle Einstellungen (die richtigen Labelfiles, die vordefinierten Labels, die farbige F_0,...) schon wie gewünscht.

2 Grundsätzliches zum Labeln

Beim Labeln von Intonation versucht man, die F_0-Kurve mithilfe eines vordefinierten Inventars (in unserem Fall die Einheiten des Stuttgarter GToBI-Modells) möglichst genau zu beschreiben. Man sollte sich hierbei auf den perzeptiven Eindruck (*Was klingt betont? Hört sich das Signal an der Stelle steigend oder fallend an?, ...*) konzentrieren. Zur Unterstützung hat man die Pitch-Kontur, an der man den Verlauf der Grundfrequenz sehen kann. Natürlich ist diese unterbrochen, da sich diese nur für stimmhafte Segmente ermitteln lässt. Außerdem kann der Algorithmus, der die Kontur erzeugt (zB bei bei *creaky voice*) eine fehlerhaft arbeiten. Vor allem einzelne, isolierte Punkte sind meistens Fehler und sollten einfach ignoriert werden.

Grundsätzlich geht man so vor, dass man die Äußerung anhört und versucht, die Intonationsphrasen und intermediären Phrasen (s.u.) festzustellen und zu markieren (im .tones-Labelfile). Danach verucht man, in den Phrasen die Akzente zu finden, indem man die prominent klingenden Silben ausmacht. Um die Art des Akzents festzulegen, muss man auf den Tonhöhenverlauf achten. Die Akzente werden in das .accents-Labelfile eingetagen.

Beim Lablen ist eine Abspiel-Funktionalität von Wavesurfer sehr geschickt: Mit `Ctrl - space` kann man das Signal zwischen zwei Labels anhöhren (wenn man vorher die entprechende Labelspur mittels eines Klicks auf die Spur aktiviert hat und den Cursor über das abzuspielende Label bewegt hat). Diese Funktionalität erleichtert einem das Ausmachen der prominenten Silben, da man jede Intonationsphrase einzeln anhören kann (indem man `Ctrl-space` über den entsprechenden Labels der .tones-Labelspur drückt). Außerdem kann man sehr schön den Tonöhen verlauf von der promineten Silbe zu der ihr folgenden ausmachen, da man durch das Abspielen der Silben (mittels `Ctrl-space` über den Labels der .syl-Labelspur) genau vergleichen kann, welche Silbe höher und welche tiefer ist.

3 Die Einheiten des Stuttgarter GToBI-Modells

3.1 Intonations-Einheiten

Es bietet sich also an, beim Labeln mit dem .tones-File anzufangen. Zuerst sollten die Intonations-Phrasen (IP), dann die intermediären Phrasengrenzen (ip) gelabelt werden.

Die Intonationseinheiten werden am Ende der Phrase, an der Grenze des letzten Lautes der Phrase gelabelt; Pausen werden als Anfang der nächsten Phrase betrachtet, die IP endet also **vor** einer Pause.

3.1.1 Intonations-Phrasen (IP)

Die IP's sind gut erkennbare, abschließende Einschnitte in der Äußerung. Die F_0-Kontur ist an dieser Stelle deutlich unterbrochen, sei es nur durch eine Pause oder durch einen nicht durch Akzente erklärbaren Verlauf der Kurve (z.b ins obere Drittel des Stimmumfangs gefolgt von einem Sprung in den mittleren Bereich). Als Label stehen für die IP's %, H%, L% und %H zur Verfügung. % markiert die "normalen" Phrasengrenzen, H% die mit steigendem Grenzton (Fragen[1]). H% zeichnet sich dadurch aus, dass die F_0-Kurve zwar ansteigt, die Silbe an der Stelle aber nicht betont klingt. %H bezeichnet einen Phrasenanfang, der extrem hoch beginnt, was sich aber nicht durch einen Akzent erklären lässt (da die Silbe nicht betont klingt). L% steht für extrem tiefe Grenztöne am Ende einer Phrase zur Verfügung, die F_0-Kurve geht dann deutlich unter die Baseline.

3.1.2 Intermediäre Phrasengrenzen (ip)

Intermediäre Phrasengrenzen werden mit "-" gekennzeichnet. Sie sind schwerer wahrnehmbar, entprechen eher einer kleinen Verzögerung, an dieser Stelle muss auch nicht unbedingt eine Pause sein. Ein Hinweis auf eine ip kann z.B. auch ein relativ starker Akzent innerhalb der Äußerung sein.

3.2 Die Pitch-Akzente

Nachdem die IP's und ip's festgelegt wurden,können die Akzente (im .accents-File) gelabelt werden. Grundsätzlich sollte hierbei darauf geachtet werden, dass

[1] allerdings **muss** nicht jede Frage unbedingt auf einem steigendem Grenzton enden. Nur wenn die F_0-Kurve deutlich über den normalen Stimmumfang hinausgeht, wird H% gelabelt.

genau in der Voakl**mitte** gelabelt wird, nicht am Ende des Vokals. Es bietet sich an, zuerst (ohne die F_0-Kontur genauer anzuschauen) den Satz anzuhören und grundsätzlich darauf zu achten, welche Silben oder Wörter (auffallend) betont sind. Dann können die einzelnen Akzente, also die Frage wie diese Silben betont sind, näher betrachtet werden. Der Verlauf der Akzente sollte die F_0-Kurve möglichst genau beschreiben.

Für die verschiedenen Akzente stehen folgende Labels zur Verfügung: H*L, L*H (die beiden grundlegenden Akzente), HH*L, L*HL und H*M (spezielle Pitchakzente). Der mit "*" gekennzeichnete Ton (der dem Stern vorausgehende) wird auf der akzentuierten Silbe realisiert, die Töne danach (trail tones) bezeichnen den weitern Verlauf.

3.2.1 fall - „Gipfelakzent"

Der Gipfelakzent wird mit dem Label H*L gekennzeichnet. Er zeichnet sich durch ein lokales F_0-Maximum in der akzentuierten Silbe gefolgt von einem steilen Pitchabfall aus. Ist die akzentuierte Silbe die letzte der Intonations-Einheit, so werden das Maximum und der Tonhöhenfall auf der gleichen Silbe (also der betonten) realisiert. Folgen der akzentuierten Silbe noch ein oder mehrere Silben, so wird das Maximum und der Anfang des Pitch-Falles auf der akzentuierten realisiert, der Abfall der Tonhöhe setzt sich auf der postakzentuierten Silbe fort.

Grundsätzlich gilt, dass sich nach H*L die F_0-Kurve in das untere Drittel des Stimmumfangs bewegt, bis die nächste Silbe bzw die Phrasengrenze erreicht ist.

3.2.2 rise

Der steigende Akzent wird mit dem Label L*H gekennzeichnet. Er ist gewissermaßen das Gegenteil des Gipfelakzents, zeichnet sich also durch ein lokales F_0-Minimum in der betonten Silbe gefolgt von einem steilen Anstieg der Tonhöhe aus. Ist die akzentuierte Silbe die letzte der Intonations-Einheit, so werden das Minimum und der Tonhöhenanstieg auf der gleichen Silbe (also der betonten) realisiert. Folgen der akzentuierten Silbe noch ein oder mehrere Silben, so wird das Minimum und der Anfang des Pitch-Anstiegs auf der akzentuierten realisiert, der Anstieg der Tonhöhe setzt sich auf der postakzentuierten Silbe fort.

Grundsätzlich gilt, dass sich nach H*L die F_0-Kurve in das obere Drittel des Stimmumfangs bewegt, bis die nächste Silbe bzw die Phrasengrenze erreicht ist.

3.2.3 early peak

Der "early-peak" wird, gemäß seinem Verlauf (high-high-low) mit dem Label HH*L gekennzeichnet. Er zeichnet sich durch ein lokales F_0-Maximum in der präakzentuierten Silbe gefolgt von einem steilen Pitchabfall aus. Der HH*L muss in mindestens zwei Silben realisiert werden, die präakzentuierte darf nicht betonbar sein.

Folgt der akzentuierten Silbe keine weitere, so wird der Fall auf ihr selbst realisiert, folgen noch weitere Silben, so beginnt der Pitchabfall auf der akzentuierten Silbe und setzt sich auf der folgenden fort. In beiden Fällen gilt aber, dass dem Pitchabfall ein lokales F_0-Maximum auf der präakzentuierten Silbe vorangehen muss.

Beim Labeln kann man meist ist der “high-high-low-Verlauf” deutlich als “Wellenlinie” in der F_0-Kurve erkennen.

3.2.4 rise-fall

Dieser Akzent wird mit dem Label L*HL gekennzeichnet. Er zeichnet sich aus durch ein lokales F_0-Minimum an der akzentuierten Silbe, gefolgt von einem lokalen F_0-Maximum und erneutem steilen Abfallen der Tonhöhe.

Wenn der akzentuierten Silbe mindestens 2 weitere Silben innerhalb der selben Intonations-Einheit folgen, wir das F_0-Minimum an der akzentuierten Silbe realisiert, der Anstieg beginnt ebenfalls an dieser Silbe, setzt sich dann aber in der nächsten fort. Der Tonhöhenabstieg beginnt in der zweiten Silbe und endet in der dritten.

Folgt der akzentuierten Silbe eine weitere, so wird das Minimum auf der akzentuierten realisiert, der Pitchanstieg beginnt ebenfalls in der akzentuierten und setzt sich auf der folgenden fort. Der Fall der Tonhöhe wird auf der postakzentuierten (also auf der zweiten Silbe) realisiert.

Ist die akzentuierten Silbe die letzte Silbe innerhalb der Intonations-Einheit, so wird der komplette Verlauf auf ebendieser Silbe realisiert.

Der L*HL ist auf jeden Fall in der F_0-Kurve als “Zacke” zu erkennen und klingt extrem betont.

3.2.5 stylized contour

Dieser Akzent wird mit dem Label H*M markiert. “M” bezeichnet hier einen Ton ungefähr in der Mitte des Sprecher-Stimmumfangs. Er tritt nur an nuklearer Position, also als letzter Akzent in der Intonations-Einheit, auf.

Gibt es postakzentuierte Silben, so folgt dem F_0-Maximum an der akzentuierten Silbe ein Levelton auf H-Niveau und die Tonhöhe fällt erst in der letztn Silbe der Intonations-Einheit ab.

Ist die akzentuierte Silbe die letzte Silbe der Phrase, o wird der Nukleus häufig verdoppelt um das F_0-Maximum dann auf seinem ersten Teil, das Mid-Target auf dem zweiten Teil realisieren zu können.

Typischerweise wird H*M beim rufen realisiert, etwa “Jö-örg!” (Verdopplung des Nukleus).

3.2.6 Downstep

Für denn fallenden Downstep-Akzent steht das diakritische Zeichen “!” zur Verfügung. Einem Downstep muss ein anderer Akzent mit F_0-Maximum (also H*L, HH*L oder H*) vorausgegangen sein. Er zeichnet sich dann dadurch aus, dass er zwar an einem lokalen Maximum der F_0-Kurve liegt, dieses aber tiefer liegt als das vorangegangene.

3.3 Linking

Beim Linking wird davon ausgegangen, dass pränukleare Akzente (also der `space`-Taste abspielenalle bis auf den letzten Akzent) der `space`-Taste abspielenihren Verlauf „aufteilen“ (bis hin zum letzten Ton vor der nächsten betonten Silbe) oder sogar ganz auslassen können. Im ersten Fall spricht man von partiellem, im zweiten Fall von komplettem Linking.

3.3.1 gelinkter H*L

Hierfür steht das Label H*/..L bzw. die beiden Teile zur Verfügung. Das lokale F_0-Maximum (H*) wird nicht unmittelbar von einem Fall der Tonhöhe gefolgt.

Handelt es sich um partielles Linking fällt die Pitch-Kontur langsam bis zur nächsten präakzentuierten Silbe ab. An dieser wird dann ..L gelabelt.

Im Falle von komplettem Linking bewegt sich die F_0-Kurve langsam auf den nächsten Akzent (H oder L) zu.

3.3.2 gelinkter L*H

Hierfür steht das Label L*/..H bzw. die beiden Teile zur Verfügung. Das lokale F_0-Minimum (L*) wird nicht unmittelbar von einem Anstieg der Tonhöhe gefolgt.

Handelt es sich um partielles Linking fällt die Pitch-Kontur langsam bis zur nächsten präakzentuierten Silbe ab. An dieser wird dann ..H gelabelt.

Im Falle von komplettem Linking bewegt sich die F_0-Kurve langsam auf den nächsten Akzent (H oder L) zu.

3.4 Unsicherheit

3.4.1 bezüglich der Art eines Labels

Ist man sich sicher, an einer bestimmten Stelle ein Label (Akzent oder IP/ip) zu benötigen, ist sich aber nur darüber unklar, um welches Label es sich handelt, so kann zu jedem Label das diakritische Zeichen ? hinzugefügt werden.

3.4.2 bezüglich dem Vorhandensein eines Akzents

Stellt sich die Frage, ob an einer bestimmten Stelle überhaupte ein Pitch-Akzent steht oder nicht, so hat man die Möglichkeit “*?” zu labeln.