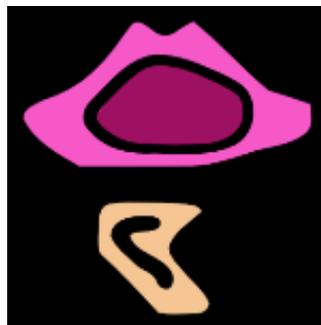
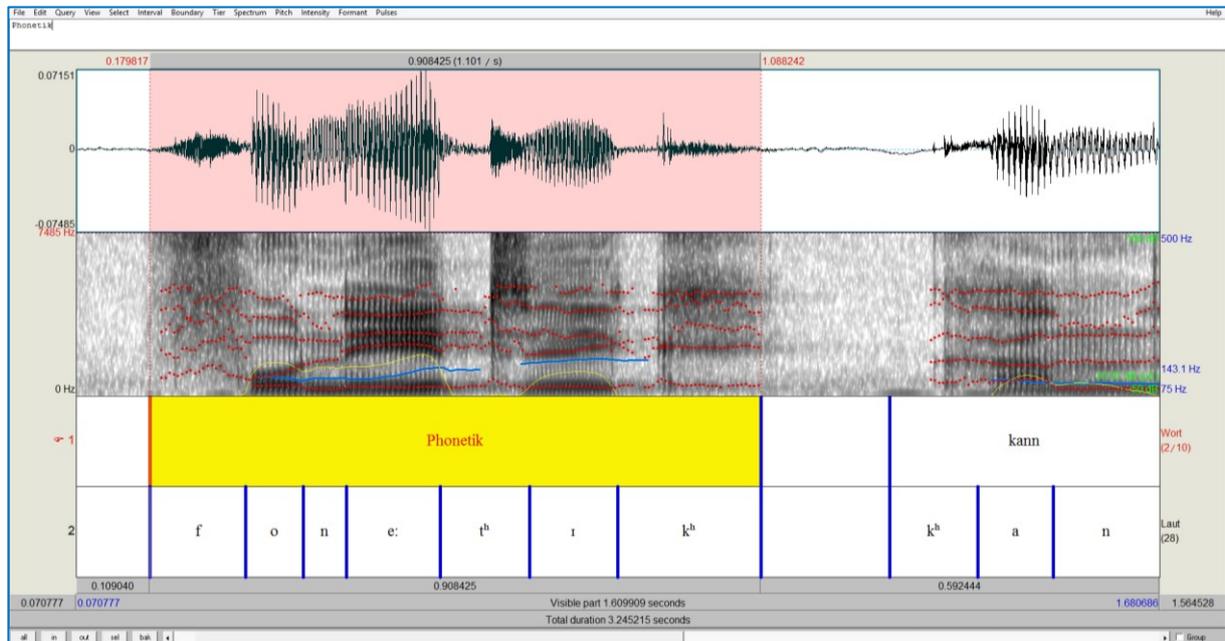


KLEINE EINFÜHRUNG IN PRAAT

Dr. François Conrad



0) EINLEITUNG

Das Programm PRAAT hat sich schnell zu *dem* Standardprogramm für phonetische Untersuchungen in der Wissenschaft entwickelt. Es ist kostenlos, äußerst performant, hat eine nur kleine Dateigröße und lässt sich daher problemlos vervielfältigen und kopieren. Einmal eingearbeitet, lassen sich alle gewünschten Analyseschritte schnell und unkompliziert durchführen. Und [Spaß](#) macht es auch noch!

Einführungen in PRAAT existieren viele. Viele der Inhalte dieser Einführung lassen sich mühelos im Netz in anderer Farbe und Form finden. Dass es mit diesem Skript jetzt noch eine weitere gibt, hat dennoch seine Gründe. Diese entspringen, wie so häufig, dem Zufall.

Im Rahmen der „Einführung in die Phonetik und Phonologie“ an der Leibniz Universität Hannover führe ich in zwei Sitzungen in die Grundelemente von PRAAT ein. Um den Einstieg möglichst „packend“ und spannend zu gestalten, verlasse ich am Beginn den Raum und die Studierenden sollen einen kurzen deutschen Satz einsprechen. Nach Rückkehr in den Seminarraum zeige ich die Grundfunktionen des Programms, indem ich diesen Satz „[stumm](#)“ höre – also indem ich das visuelle Signal anschau und die einzelnen Laute bestimme, dabei aber nichts höre. Meistens gelingt es (zum Glück)! Aus diesem Einstiegs spiel heraus wurde mir deutlich, dass es keine zusammenfassende Beschreibung des spektralen Bildes der einzelnen deutschen Laute gibt. Ein solches kann jedoch gerade am Beginn der Beziehung mit PRAAT überaus hilfreich sein, wird hierbei doch die technische – „was muss ich wie klicken, damit dieses oder jenes ausgeführt wird“ – sowie die auditive – „ich höre ein [a] im Wort ‚machen‘ – Handhabe von Anfang an um einen konkreten visuellen Eindruck ergänzt. Folgende Einführung führt somit als [Herzstück](#) ein [Inventar der akustischen Visualisierung deutscher Laute](#) zum schnellen Nachschlagen und Vergleichen, das meines Wissens bislang im Zusammenhang mit PRAAT gefehlt hat. Ergänzt wird dieses Inventar um einige der klassischen Funktionen des Programms wie der Annotation und Extrahierung von Lauten und Wörtern sowie der Aufnahme eigener Sprachsignale.

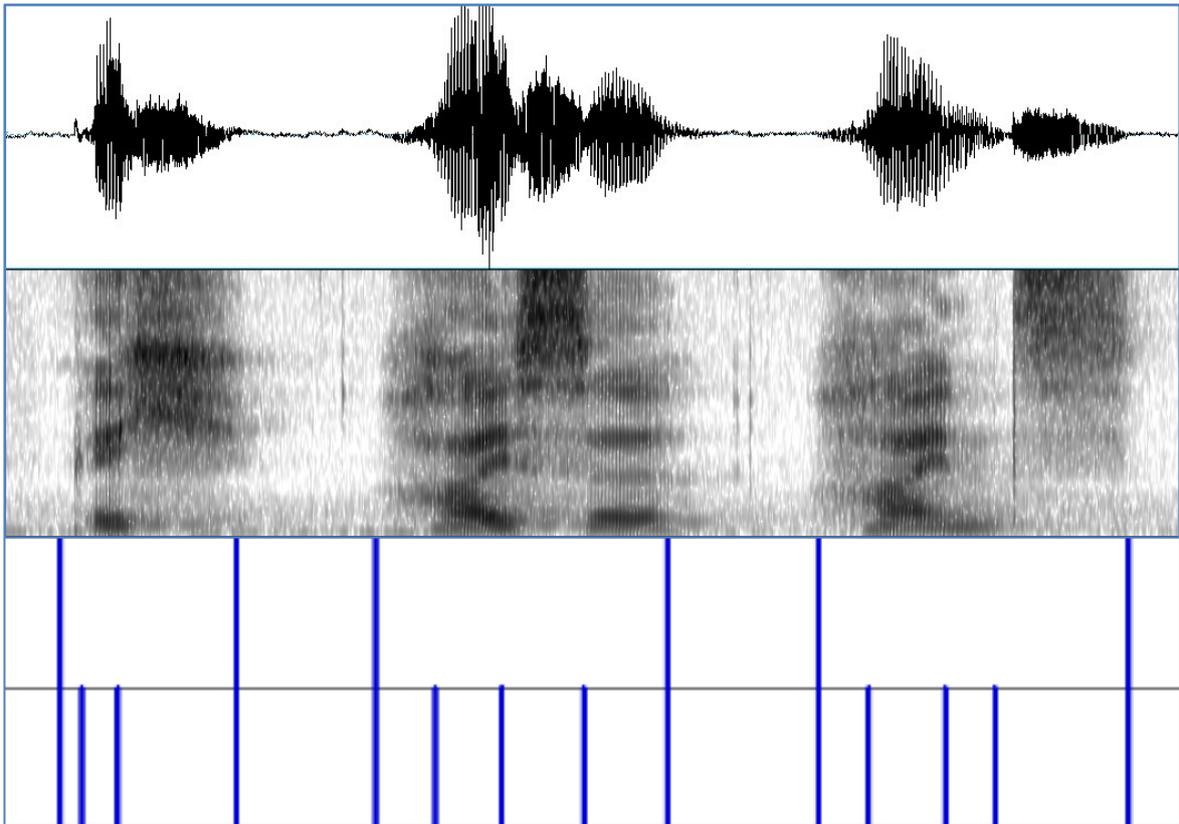
Das Skript entstand im Rahmen eines Blockseminars an der Leibniz Universität Hannover („Einführung in das phonetische Programm PRAAT“). In zwei Tagen werden den interessierten Studierenden vor allem diejenigen Funktionen beigebracht, die sie im Laufe ihres Studiums sowie im späteren Berufsleben besonders gebrauchen können und die zudem nach meiner Einschätzung besonders spannend sind. Aus diesem Grund enthält die Einführung auch eine Anleitung zum Erstellen von [Vokaltrapezen](#) sowie zum Messen der ‚voice onset time‘ (VOT). Beides können die Studierenden im Laufe des Blockseminars an ihrer eigenen Sprache sowie derjenigen der KommilitonInnen ausprobieren.

Ich hoffe, dass diese Einführung die bereits im Netz existierenden sinnvoll ergänzt, den Zugang zu der eigenen Sprachbeschreibung erleichtert – und v.a. auch die Freude an PRAAT verstärkt. In diesem Sinne schalte ich das oben skizzierte [Spiel](#) vor: Die folgende Darstellung zeigt das Spektrogramm eines kleinen deutschen Satzes. Es beinhaltet ein finites Verb sowie einen weiblichen deutschen Vornamen. Als Hilfestellung sind die Anzahl an Wörtern und Lauten vermerkt. Mein Wunsch und meine Hoffnung sind, dass nach der Lektüre des Skripts – und natürlich dem ergänzenden Ausprobieren – der Satz „stumm“ gelesen werden kann. Die Auflösung findet sich am Ende der Einführung ...

Viel Spaß bei der Lektüre und ich freue mich über alle möglichen Rückmeldungen!

[Dr. François Conrad](#)
im April 2019

francois.conrad@germanistik.uni-hannover.de



Inhaltsverzeichnis

- 0) Einleitung1**
- 1) Erste Schritte4**
 - 1. Allgemein4
 - 2. Was kann ich mit dem Programm (unter anderem) machen?4
 - 3. Installieren4
 - 4. Wie öffne ich eine Sound-File?.....4
 - 5. Wie spiele ich eine Sound-Datei ab?5
 - 6. Selber etwas aufnehmen5
 - 7. Wie beschrifte ich einen Sound? (Segmentierung und Transkription).....5
 - 8. Wie extrahiere ich einen einzelnen Laut/ein einzelnes Wort?6
 - 9. Wie speichere ich?.....6
 - 10. Wichtige Praat-Shortcuts (die viel Zeit und Nerven sparen!)7
- 2) Das Analysefenster7**

3) Kürzel der für das Deutsche wichtigen phonetischen Symbole	10
1. Konsonanten	10
2. Vokale	11
4) Wie sehen die (prototypischen) deutschen Laute im Spektrogramm aus?	12
A) Konsonanten	12
1. Grundsätzliches	12
2. Plosive [p, t, k] und [b, d, g] (und [ʔ])	13
3. Frikative [f, s, ʃ, ç, x, h] und [v, z, ʒ].....	15
4. Liquide [l, ʁ, r, R]	18
5. Nasale [n, m, ŋ]	20
6. Approximant [j]	21
B) Vokale und Diphthonge	22
1. Vokale	22
2. Diphthonge	23
5) Vokaltrapeze	25
1. Vorbemerkungen.....	25
2. Formanten messen.....	25
3. Excel-Tabelle anlegen	26
4. Vokaltrapez erstellen.....	26
5. Normalisierung der Werte	27
6. Komplexe Vokaltrapeze	28
7. Darstellung von Diphthongen	29
8. Literatur	29
6) Voice Onset Time (VOT).....	30
7) Anleitungen, Online-Materialien etc.	32
8) Auflösung der „stummen“ Analyse.....	33

1) ERSTE SCHRITTE

1. ALLGEMEIN

- entwickelt von Paul [BOERSMA](#) und David [WEENINK](#) von der Universität Amsterdam
- wird von zahllosen WissenschaftlerInnen in der ganzen Welt benutzt

2. WAS KANN ICH MIT DEM PROGRAMM (UNTER ANDEREM) MACHEN?

- 1) [phonetische Analysen](#): einzelne Laute (Qualität, Dauer, ...), prosodische Phänomene (Intonation, Zusammenspiel zwischen den Lauten, etwa Sandhi-Phänomene, ...)
- 2) [Transkription](#): Wörter/Laute Buchstaben zuordnen (und besser wiederfinden)
- 3) Sprachdaten und Laute [verändern](#)
- 4) [künstliche](#) Geräusche, Stimmen, Laute erstellen
- 5) all diese Analysen auch [graphisch darstellen](#)

Was habe ich davon?

- a) ich kann gesprochene Sprache wissenschaftlich [analysieren](#)
- b) ich kann [Variation](#) beim gleichen und bei verschiedenen SprecherInnen feststellen
- c) ich kann theoretische Annahmen und Überlegungen [empirisch](#) untersuchen

3. INSTALLIEREN

- 1) Herunterladen von PRAAT:
 - „praat“ googeln, oder direkt über
 - <http://www.fon.hum.uva.nl/praat/>
- 2) Kleine Datei, kann kopiert werden auf einen USB-Stick, auf andere Laptops, ...
- 3) Läuft auf allen Betriebssystemen

4. WIE ÖFFNE ICH EINE SOUND-FILE?

- 1) im Fenster "Praat objects" wähle ich oben „[Open](#)“
- 2) ich drücke auf „[Read from file...](#)“ (für normale Sound-Files) oder „[Open long sound file...](#)“ (für längere Sound-Files > bei langen Aufnahmen wird die Arbeitsgeschwindigkeit stark reduziert, wenn diese normal geöffnet werden)
- 3) diese File erscheint dann bei "[Objects](#)". Ich klicke sie an und wähle dann rechts „[View & Edit](#)“ – oder ich drücke die [ENTER](#)-Taste. Ein neues Fenster öffnet sich, wo die Sound-File als Oszillogramm und Spektrogramm sichtbar wird

5. WIE SPIELE ICH EINE SOUND-DATEI AB?

Unterhalb des Spektrogramms (und gegebenenfalls unterhalb der annotierten Wörter und Laute) befinden sich **drei graue Balken**. Ein Klick in einen dieser Balken lässt verschiedene Bereiche abspielen:

- 1) Der **obere** der drei Balken gibt nur den ausgewählten Bereich wieder (im Oszillogramm rosa, im TextGrid gelb unterlegt). Dieser Balken erscheint ebenfalls über dem Oszillogramm und führt bei Klick zum selben Ergebnis.
- 2) Der **mittlere** Balken gibt den Sound vom gesamten *sichtbaren* Analysefenster wieder.
- 3) Der **untere** Balken spielt die *gesamte* Sound-File ab, auch die nicht-sichtbaren Teile vor und nach dem ausgewählten Bereich.

Durch Klick auf die **ESCAPE**-Taste wird das Vorspielen gestoppt/abgebrochen.

Alternativ kann durch Klick auf die **TAB**-Taste (unterhalb der ESC-Taste) der ausgewählte Bereich abgespielt werden. Insbesondere bei häufiger Wiederholung des Abspielens kann die TAB-Taste viel Zeit sparen, wenn die Hände ohnehin auf der Tastatur aufliegen (und nicht auf der Maus), etwa beim Annotieren.

6. SELBER ETWAS AUFNEHMEN

- 1) im Fenster „Praat objects“ wähle ich oben „**New**“
- 2) ich drücke auf „**Record mono Sound**“ (bei Stereo erscheinen 2 Oszillogramme, was unnötig ist)
- 3) ein neues Fenster geht auf
- 4) rechts unten gebe ich der Aufnahme einen Namen (sonst heißt sie „untitled“)
- 5) ich drücke unten links auf „**Record**“ und rede
- 6) wenn ich fertig bin, drücke ich auf „**Stop**“
- 7) ich drücke unten rechts auf „**Save to list & close**“
- 8) das, was ich gerade gesprochen habe, erscheint als Sound-File im „**Praat Objects**“-Fenster

7. WIE BESCHRIFTE ICH EINEN SOUND? (SEGMENTIERUNG UND TRANSKRIPTION)

- 1) ich klicke die Sound-File an, die ich beschriften will
- 2) ich klicke in der rechten Leiste auf „**Annotate**“
- 3) ich klicke auf „**To TextGrid...**“. Ein neues Fenster öffnet sich.
- 4) bei „**All tier names**“ gebe ich einen Namen für die einzelnen Schichten (=tier) an (z.B. **Satz Wort Laut**). Die einzelnen Schichten können einzeln beschriftet werden (in die erste Schicht würde man etwa den Satz schreiben, in die zweite die einzelnen Wörter, in die dritte die einzelnen Laute). Die einzelnen Namen werden hier durch ein Leerzeichen voneinander getrennt. Bei „**Which of these are point tiers**“ lösche ich „bell“ und lasse das Fenster leer. Ich drücke auf „**ok**“
- 5) Im „**Praat objects**“-Fenster erscheint ein neues Objekt, dessen Name mit „**TextGrid**“ beginnt

- 6) ich klicke beide Objekte (**Sound und TextGrid**) an, dabei halte ich ‚Ctrl‘ gedrückt (oder ich fahre mit der Maus über beide, während ich die linke Maustaste gedrückt halte)
- 7) ich drücke rechts **„View & Edit“** (oder die ENTER-Taste). Ein neues Fenster öffnet sich.
- 8) Die verschiedenen Leisten, deren Name rechts erscheint, können nun beschriftet werden. Es wird jeweils die Leiste beschriftet, die gelb erscheint. Um einzelne Sätze/Wörter/Laute voneinander zu trennen, lassen sich Trennstriche (**„Boundaries“**) zeichnen. Diese werden automatisch gesetzt, wenn in den Kreis oberhalb der Striche geklickt oder die ENTER-Taste gedrückt wird. Die dadurch entstandenen Unterteilungen lassen sich individuell beschriften und manuell neu anpassen, indem durch Gedrückthalten der linken Maustaste der Strich verschoben wird.
- 9) Für **phonetische Symbole**: sie werden mit einer Kombination **„/xy“** verschriftet. Bei **„Help“** im **„Praat object“**-Fenster ganz oben rechts erscheint **„Phonetic symbols“**. Darüber gelange ich zu Listen, wo die Kombinationen für die einzelnen Laute stehen, die ich dann einfach im TextGrid eintrage. Sie erscheinen unten als phonetische Symbole, in der Leiste ganz oben in der Kombination **„/xy“** (siehe hierzu ebenfalls Kapitel 3).

8. WIE EXTRAHIERE ICH EINEN EINZELNEN LAUT/EIN EINZELNES WORT?

- 1) ich **selektiere** mit der Maus einen Teil (dieser erscheint **rot**)
- 2) ich klicke ganz oben links auf **„File“**
- 3) ich klicke auf **„Extract selected sound (preserve times)“**
- 4) dieser Sound ist jetzt als einzelnes Objekt im **„Praat object“**-Fenster; ganz unten kann ich ihn mit **„Rename...“** umbenennen
- 5) wenn ich den Sound schon beschriftet habe, kann ich mit **„Extract selected TextGrid (preserve times)“** auch das entsprechende TextGrid extrahieren. Die zwei Objekte kann ich wieder zusammen öffnen und habe dann nur den Sound mit Beschriftung, den ich gerne hätte

9. WIE SPEICHERE ICH?

- 1) jedes Objekt muss einzeln gespeichert werden. Wenn ich PRAAT schließe, ohne gespeichert zu haben, sind diese Objekte verloren!
- 2) ich klicke im **„Praat object“**-Fenster das Objekt an, das ich speichern will
- 3) ich klicke oben auf **„Save“**
- 4) wenn es eine Sound-File ist > **„Save as WAV file...“**; Vorteil: ich kann sie auch mit anderen Programmen öffnen.
- 5) wenn es eine **TextGrid-File** ist: ich klicke auf **„Save as text file...“**
- 6) ich kann auch beides zusammen speichern: > **„Save as binary file...“**; Vorteil: wenn ich diese Datei öffne, habe ich Sound und Text gemeinsam gleich in PRAAT geöffnet; Nachteil: sie lässt sich auch nur in PRAAT öffnen
- 7) Wenn ich die Dateien später öffne, erscheinen sie direkt im **„Praat objects“**-Fenster

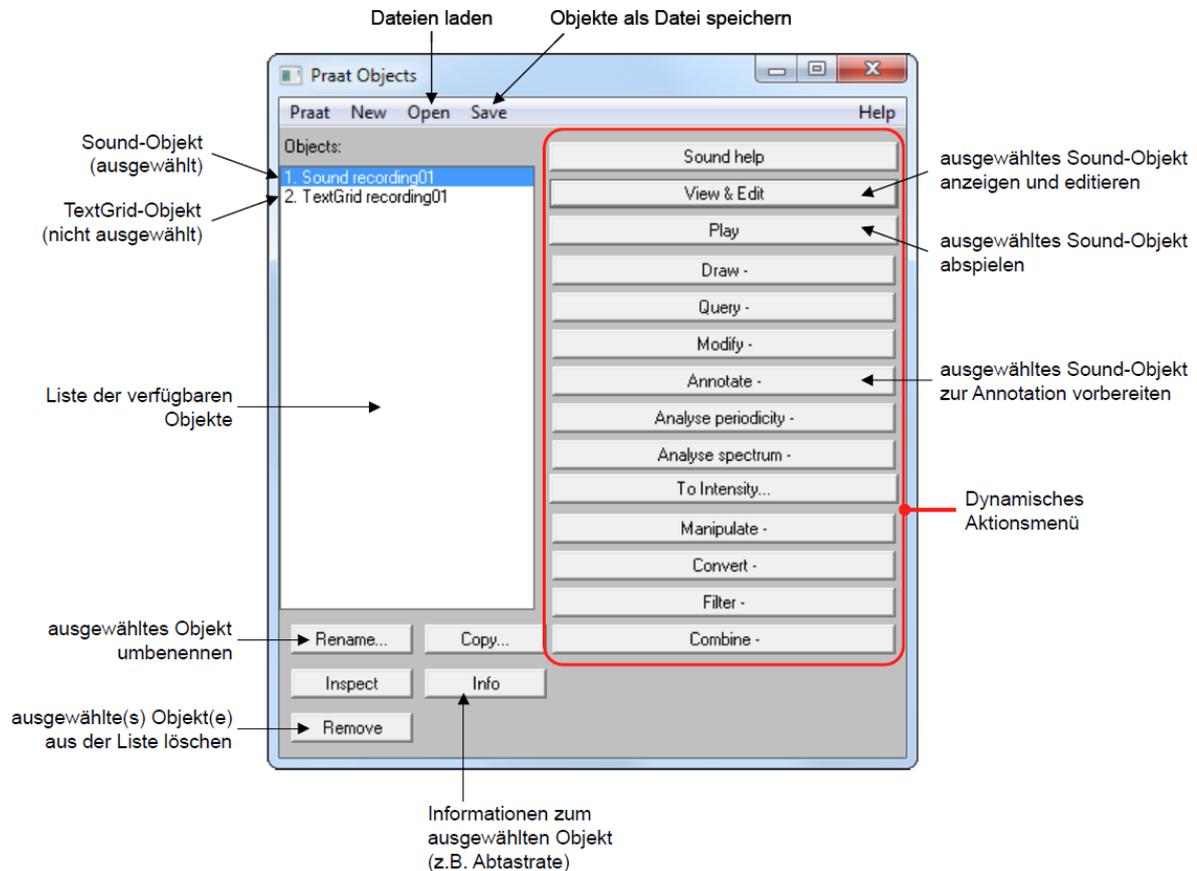
10. WICHTIGE PRAAT-SHORTCUTS (DIE VIEL ZEIT UND NERVEN SPAREN!)

Strg + i	reinzoomen (> in)
Strg + o	rauszoomen (> out)
Strg + n	zoomen auf ausgewählten Bereich (dieser erscheint immer gelb)
Strg + a	zeigt wieder das gesamt Spektrogramm an
Strg + s	speichern
alt + ←/→	nächstes Intervall wählen
alt + löschen (←)	Boundary löschen
TAB	abspielen
ESC	abspielen stoppen

P.S.: Dies sind die Shortcuts für **Windows**, für Apple-Computer weichen Sie leicht ab!

2) DAS ANALYSEFENSTER

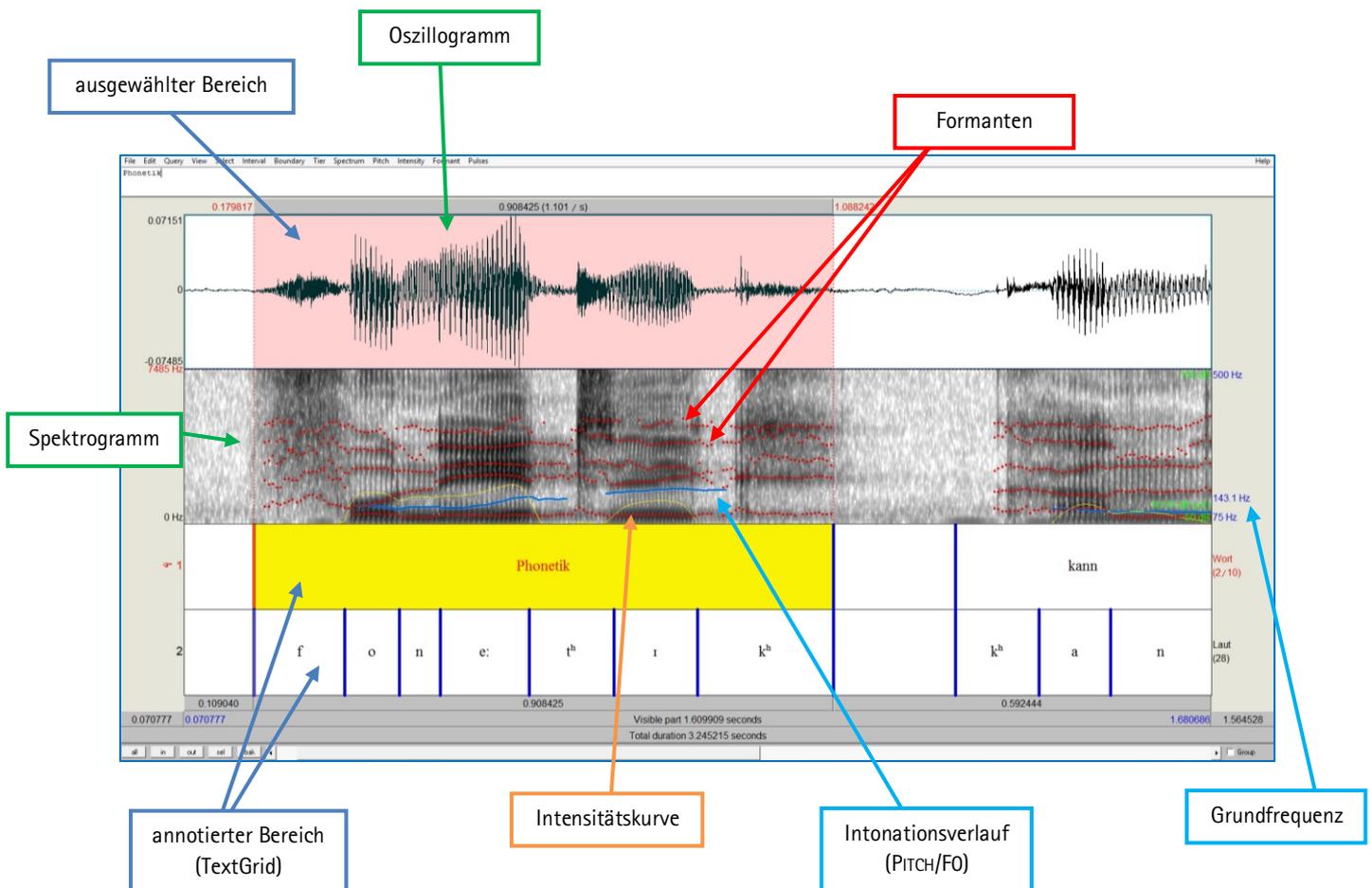
Die einzelnen **Optionen** in dem „Praat objects“-Fenster gibt folgende (selbsterklärende) Übersicht wieder (aus: MAYER, Jörg: *Phonetische Analysen mit Praat. Ein Handbuch für Ein- und Umsteiger*, S. 25 – online verfügbar unter www.praatpfanne.lingphon.net):



Das folgende Beispielfenster zeigt einen Auszug aus dem Satz „*Phonetik kann sehr viel Spaß machen.*“ Die einzelnen Begriffe werden über bzw. unter der Darstellung näher erläutert.

OSZILLOGRAMM: Oberhalb des Spektrogramms wird das **Oszillogramm** angezeigt. Es zeigt für ein Sprachsignal die Veränderung des Schalldrucks über die Zeit und ist vielen LeserInnen sicherlich aus Musikbearbeitungsprogrammen bekannt. Für unsere Analysen ist das Oszillogramm vor allem bei der Segmentierung von Lauten hilfreich (Vokale unterscheiden sich durch höhere und regelmäßige Ausschläge/“Striche“ von einer Vielzahl von Konsonanten, die als reine Geräusche eher chaotische Haufen von Schwärzungen darstellen, vgl. die beiden ersten Laute [f] und [o]) sowie bei der Erfassung der VOICE ONSET TIME (siehe Punkt 6).

SPEKTROGRAMM: Das wichtigste Analysefenster in PRAAT ist das **Spektrogramm** (Synonym: **Sonagramm**). Ein Spektrogramm visualisiert die Frequenzen eines Sprachsignals (auf der X-Achse) im Verlauf der Zeit (Y-Achse), indem es sie als Schwärzungen abbildet. Je schwärzer die abgebildeten Frequenzen, desto mehr spektrale Energie weisen sie auf. Jeder Sprachlaut hat ein typisches Schwärzungsmuster, das sich aber natürlich von Person zu Person (zum Teil auch von Sprache zu Sprache) leicht unterscheidet.



GRUNDTONVERLAUF (blaue Kurve): Die blauen Punkte repräsentieren die Schwingungen der Stimmlippen (bei stimmhaften Lauten). Als Linie zeigen sie den **Tonhöhenverlauf (PITCH, F0)** an, was für Untersuchungen der **Intonation** wichtig ist. Z.B. lassen sich Aussage- von Fragesätzen unterscheiden: Bei Fragesätzen im Deutschen steigt die Intonationskurve in der Regel gegen Ende der Äußerung an, während sie in Aussagesätzen abfällt.

Der **durchschnittliche Grundton (Grundfrequenz)** in einem ausgewählten Ausschnitt wird an der rechten Seite angezeigt. Hierdurch kann schnell erkannt werden, ob es sich bei einem Sprecher um einen **Mann** oder eine **Frau** (oder ein Kind) handelt. Die Grundfrequenz liegt bei Frauen in der Regel zwischen 150–350 Hz, bei Männern zwischen 80–200 Hz.

INTENSITÄT (gelbe Kurve): Repräsentiert die wahrgenommene Lautstärke in dB (Dezibel). Hierdurch lassen sich zum Beispiel betonte von unbetonten Silben unterscheiden (die betonten Silbenkerne, in der Regel Vokale, zeigen einen höheren „Ausschlag“ als unbetonte Silbenkerne).

FORMANTEN (rote Punkte): Formanten stellen Frequenzen mit erhöhter Energie dar (ausführlicher unter Punkt 4 B). Diese sind v.a. für die Bestimmung und Analyse von Vokalen von entscheidender Bedeutung (und gelegentlich, insbesondere aufgrund von Formantbewegungen – so genannten **Transitionen** –, auch für Konsonanten).

Alle diese Darstellungselemente können durch **Klick in der obersten Leiste** (unter dem jeweiligen Namen) ausgeschaltet werden. Insbesondere die Darstellung der Formanten (aber auch des Pitches) erfordert einen großen Rechenaufwand und verlangsamt die Darstellung. Es empfiehlt sich daher, diese **nur auszuwählen** und zu zeigen, **wenn explizit damit gearbeitet wird**.

2. VOKALE

	front	central	back
close	i y i y	ɨ ʉ li- lu-	ɯ u \mt u
close centralized	ɪ ʏ lic lyc		ʊ \hs
close-mid	e ø e lo/	ə ɘ le- lo-	ɤ ɔ \rh o
		ɚ \sw	
open-mid	ɛ œ \ef \oe	ɜ ɞ \er \kb	ʌ ɔ \vt \ct
	æ \ae	ɚ \at	
open	a ɶ a \Oe		ɑ ɒ \as \ab

Zusätzlich zu diesen Grundsymbolen können **Diakritika** (also Symbole über, unter oder neben den Hauptsymbolen) hinzugefügt werden. Für das Deutsche relevant sind etwa:

h – für die aspirierten Plosive (bei t^h würde man entsprechend „t^h“ eingeben)

$\text{\textsubscript{v}}$ – für silbische Konsonanten (bei Schwa-Elision, etwa η bei *reden* < „n_v“)

Eine Liste der Diakritika findet sich unter: „HELP“ > „PHONETIC SYMBOLS: **DIACRITICS**“

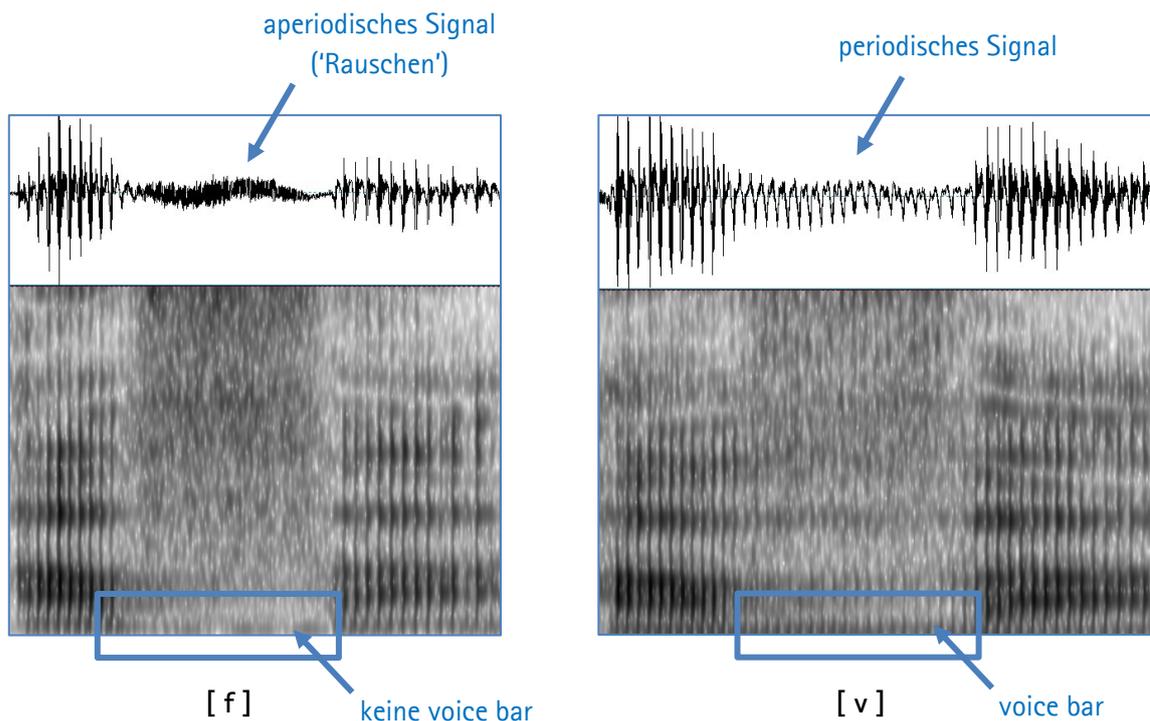
4) WIE SEHEN DIE (PROTOTYPISCHEN) DEUTSCHEN LAUTE IM SPEKTROGRAMM AUS?

A) KONSONANTEN

Im Folgenden werden prototypische Realisierungen der Konsonanten des (Standard)Deutschen mit ihrer Darstellung im Sonagramm vorgestellt. Die Konsonanten wurden jeweils im Kontext zwischen /a/ geäußert, also [a_a]. Das Fenster in PRAAT zeigt Frequenzen bis 7500 Hz., die Einstellung, mit der ich am liebsten arbeite (Sprecher: F.C.; Abweichungen sind entsprechend angegeben.)

1. GRUNDSÄTZLICHES

- 1) *Stimmhafte* Laute lassen sich von *stimmlosen* unter anderem durch die so genannte ‚voice bar‘ (Stimmbalken im untersten Frequenzbereich, unterhalb 500 Hz) unterscheiden (Beispiel: stimmloser/stimmhafter labiodentaler Frikativ [f] vs. [v]). Außerdem stellen die stimmlosen Frikative Geräusche dar, was sich im Oszillogramm in aperiodischem (zufälligem) Rauschen äußert. Stimmhafte Frikative hingegen zeigen im Oszillogramm periodische, also regelmäßige Muster auf.



2. PLOSIVE [p, t, k] UND [b, d, g] (UND [ʔ])

Bei den Plosiven werden **drei Phasen** unterschieden (Details auch unter Kapitel 6):

a. **Verschlussphase:**

Der Mundraum ist komplett verschlossen, entsprechend zeigen sich keine Schwärzungen im Spektrogramm (ausgenommen häufig voice bar bei stimmhaften Plosiven)

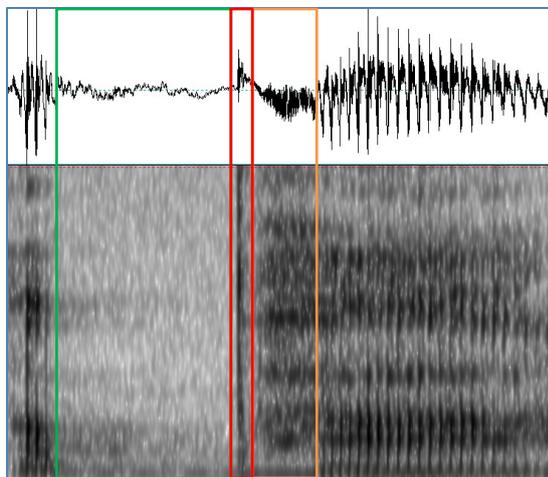
b. **Verschlusslösung** (,burst'):

Der Verschluss wird explosionsartig gelöst. Dies zeigt sich in der Regel durch einen oder mehrere dicke schwarze Striche über den gesamten gezeigten Spektralbereich.

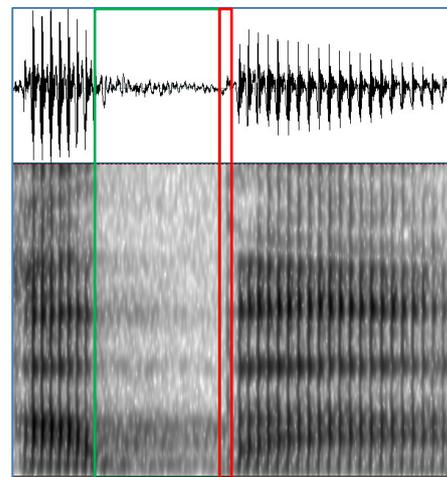
c. **Aspirationsphase:**

Bei den *stimmlosen* Plosiven (im Deutschen und anderen Sprachen) wird vor Einsetzen des Vokals ein [h] gebildet, in dem ein großer Anteil Restluft bei der Explosion „ausgepustet“ (aspiriert > „Aspiration“) wird (daher die enge Transkription [p^h], [t^h], [k^h]).

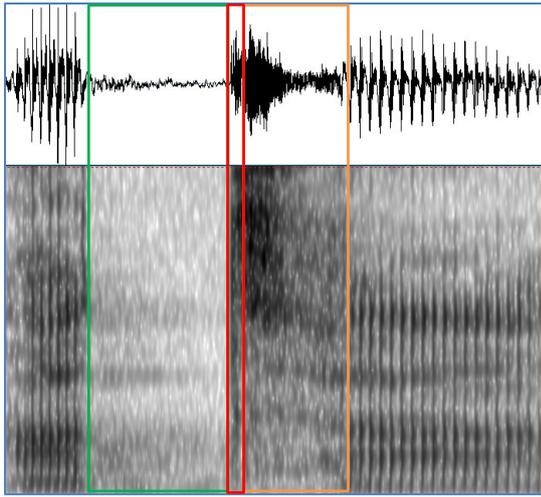
Bei den *stimmhaften* Plosiven bleibt die Behauchungsphase aus (bzw. sie ist sehr kurz und kaum zu erkennen), der Konsonant geht direkt in den Vokal über. Daher (und aufgrund der Energie, die für die Schwingung der Stimmbänder = Stimmhaftigkeit verwendet wird) sind die stimmhaften Plosive in der Regel kürzer.



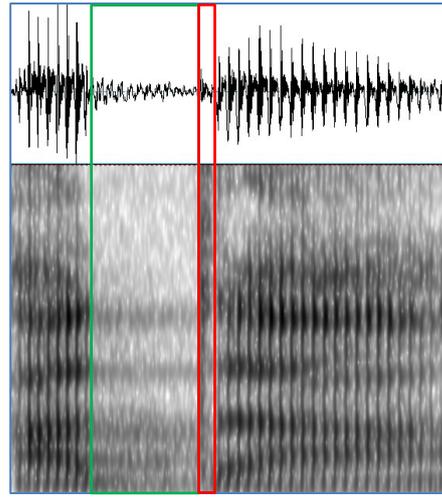
[p]



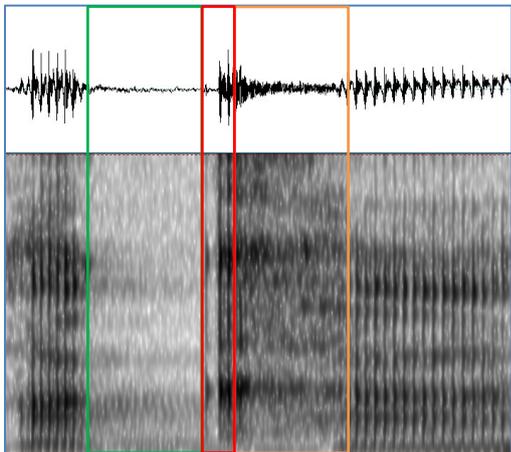
[b]



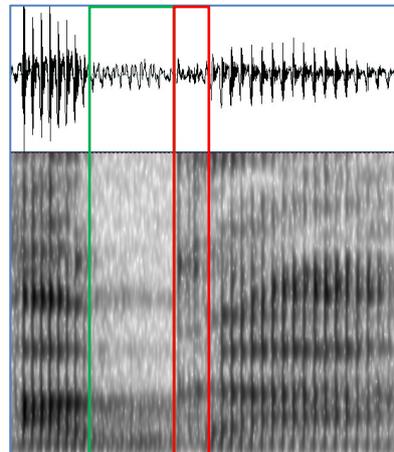
[t]



[d]

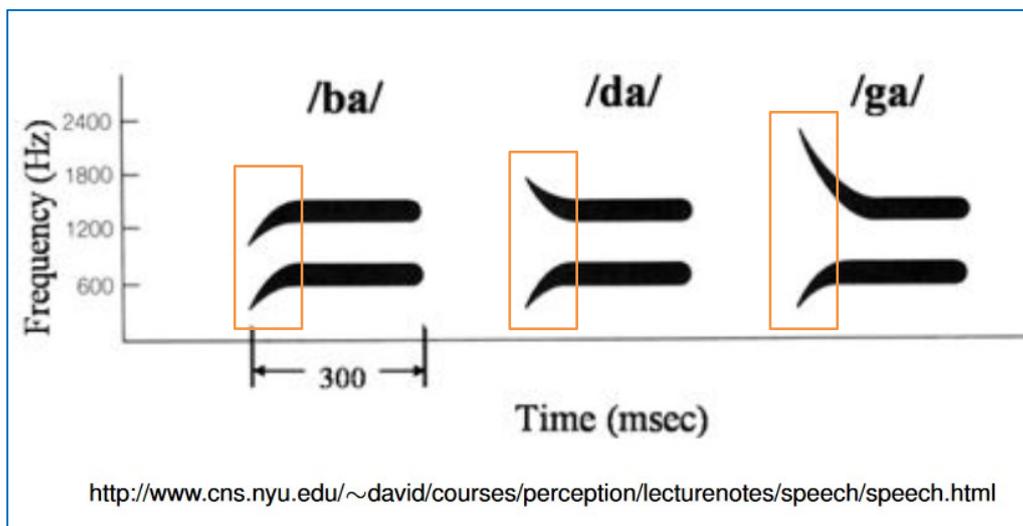


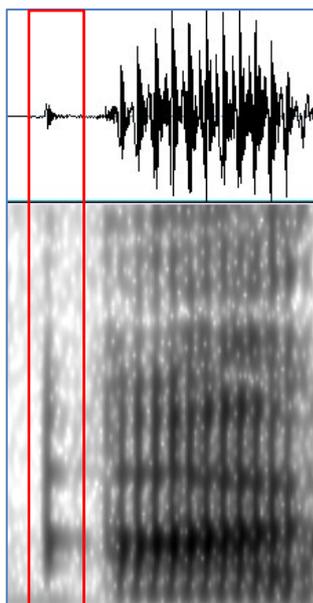
[k]



[g]

Zur Unterscheidung der stimmhaften Plosive werden auch die **Formanttransitionen** verwendet, nach dem Schema in der folgenden Abbildung:





[ʔ a]

Der **glottale Plosiv** (Glottisverschlusslaut, ‚Knacklaut‘) ist ein in der Regel einmaliges (gelegentlich zwei, manchmal stattdessen auch ein „Knarren“ aus sehr vielen Knacklauten) Aufgehen und Schließen der Stimmbänder. Im Deutschen kommt er in 2 Positionen vor:

- Im vokalischen Wortanlaut (betont oder unbetont): *ʔein, ʔana'log*
- innerhalb eines Wortes vor betonter Silbe, die mit Vokal anlautet: *cha'ʔotisch, wo'ʔanders*

Im Spektrogramm zeigt er sich entsprechend durch **eine vertikale, durchgezogene Linie** (manchmal mehrere) **vor Beginn des Vokals**. Im Oszillogramm ist er leicht abgesetzt vor dem eigentlichen Vokal als kleiner „Knubbel“ zu erkennen.

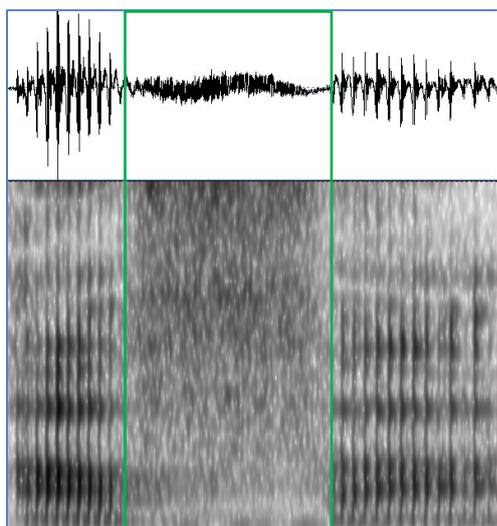
In der gesprochenen (Alltags-)Sprache tritt der Knacklaut aus unter anderem koartikulatorischen Gründen häufig nicht auf und/oder ist im Spektrogramm nicht oder kaum sichtbar. Auditiv ist er fast nur zu hören, wenn er isoliert in PRAAT angehört wird – in Verbindung mit dem folgenden Vokal geht er auditiv oft „verloren“.

3. FRIKATIVE [f, s, ʃ, ç, x, h] UND [v, z, ʒ]

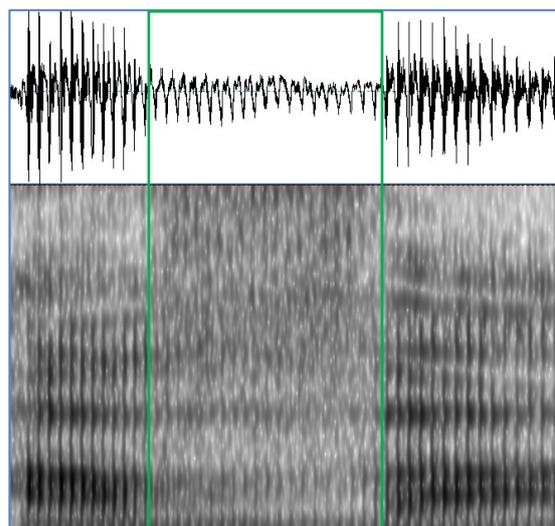
Stimmlose Frikative sind (mit Ausnahme von [x] und [h]) klar von den umgebenden Lauten abgegrenzt. Bei den stimmhaften Frikativen ist die ‚voice bar‘ erkennbar und die Vokalformanten greifen noch in den Frikativ hinein (letzteres ist auch bei [x] und [h] der Fall). Durch die Stimmhaftigkeit unterscheiden sich die Frikativpaare auch klar im Oszillogramm (mit klaren, periodischen Schwingungen bei den stimmhaften Frikativen). (Punkte 3–5: Sprecherin Mitte 30 aus Hannover)

1) labiodentaler Frikativ

An den Zähnen gerät die Luft in starke Turbulenzen. Dies zeigt sich in den fast gleichmäßigen grauen Musterungen (ohne klaren Schwerpunkt) über das gesamte Fenster.



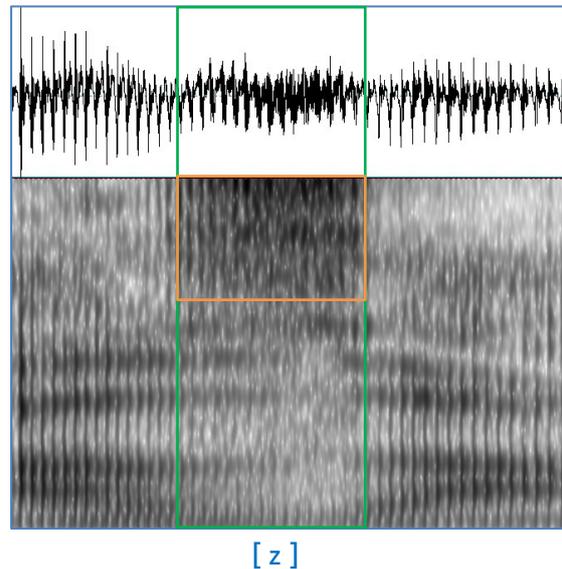
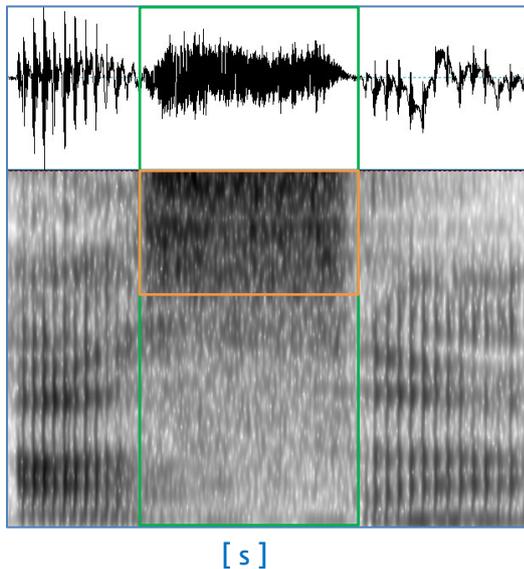
[f]



[v]

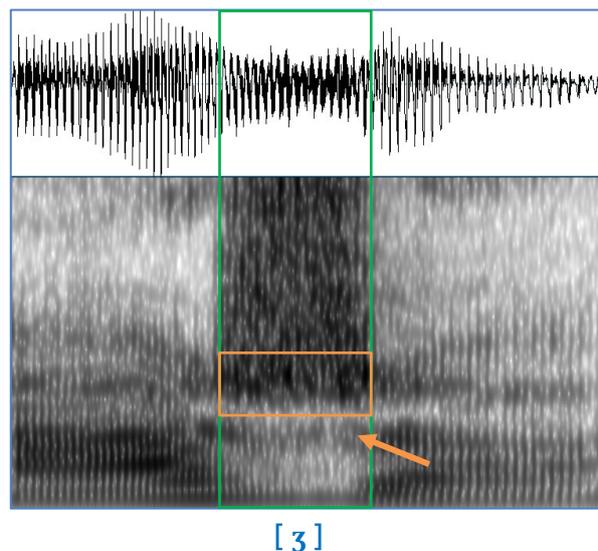
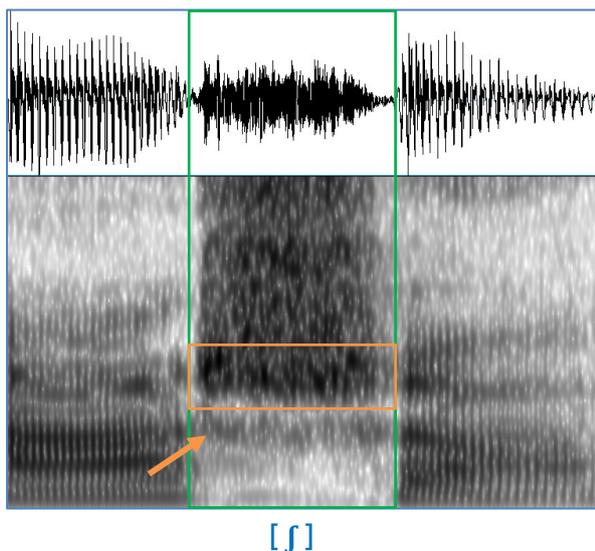
2) alveolarer Frikativ

Die alveolaren Frikative sind leicht erkennbar an ihren starken Schwärzungen **im hohen Frequenzbereich**, während im mittleren und vor allem im unteren Frequenzbereich kaum Schwärzungen zu erkennen sind.



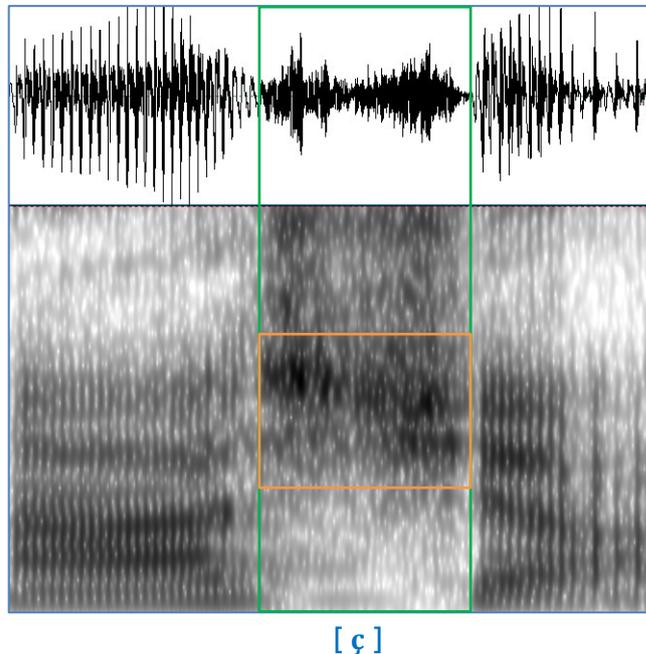
3) post-alveolarer Frikativ

Bei den post-alveolaren Frikativen liegt ein (manchmal zwei) **besonders dunkler Balken** (=hohe Energiekonzentration) etwa auf der Höhe von einem Drittel des Frequenzbereichs. Auch darüber, bis in die hohen Frequenzbereiche, zeigt sich eine hohe Energiedichte. Unterhalb des dunklen Balkens befindet sich häufig noch ein weiterer, hellerer (**Pfeil**). Dieser zeigt die bei älteren SprecherInnen häufig starke, bei jüngeren SprecherInnen in der Regel nur noch leichte Hervorstülpung der Lippen an.



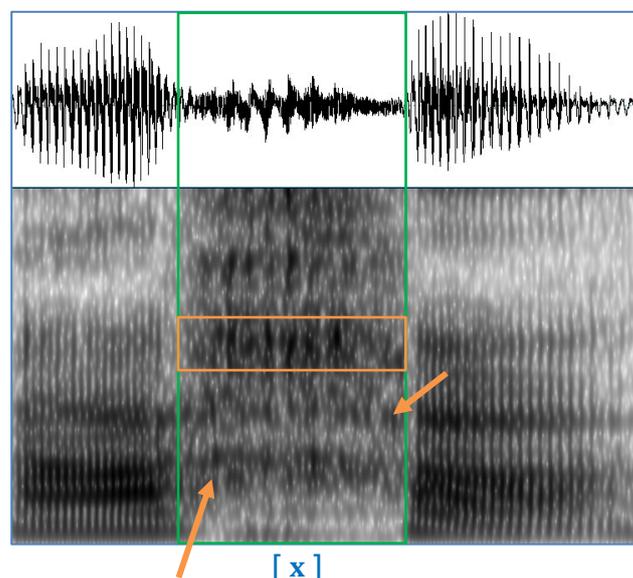
4) stimmloser palataler Frikativ

Der palatale Frikativ („*ich*-Laut“, im Standarddeutschen nur stimmlos) sieht dem post-alveolaren Frikativ recht ähnlich. Der (manchmal die) **dunkle Balken** liegt jedoch höher (etwa in der Mitte des Frequenzbereichs), was die vergleichsweise Rückverlagerung der Zunge anzeigt (post-alveolar > palatal).



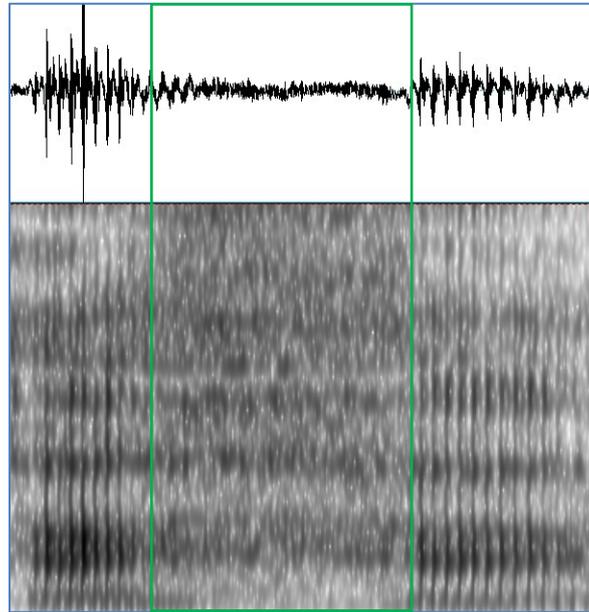
5) stimmloser velarer Frikativ

Je weiter hinten der Frikativ realisiert wird, umso höher liegt der Bereich mit der höchsten Energiekonzentration (= **dunkler Balken**). Beim velaren Frikativ liegt er entsprechend noch etwas höher als beim palatalen Frikativ, ist in der Regel aber schmaler und schwerer zu erkennen. Der velare Frikativ macht insgesamt einen eher „zerstückelten“ Eindruck. Dies geht unter anderem darauf zurück, dass anders als beim palatalen Frikativ aus koartikulatorischen Gründen klare Formantstrukturen von den benachbarten Vokalen auftauchen (**Pfeile**).



6) stimmloser glottaler Frikativ

Der stimmlose, glottale Frikativ [h] ist häufig schwer zu erkennen. Er besteht aus praktisch ungehindert ausströmender Luft und hat daher in der visuellen Darstellung im Spektrogramm die Tendenz, die Muster und Strukturen der ihn umgebenden Laute anzunehmen.

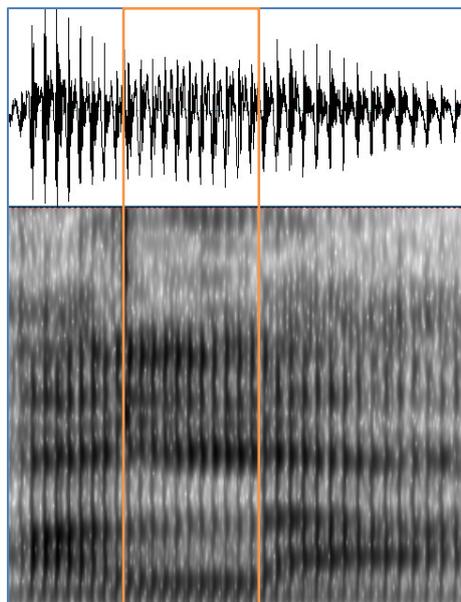


[h]

4. LIQUIDE [l, ʎ, r, R]

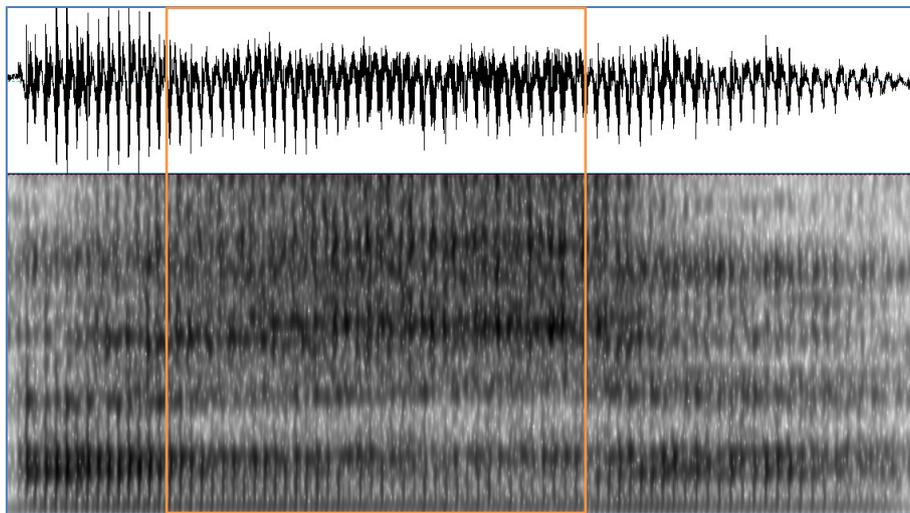
Die Liquide sind aufgrund ihrer hohen Stimmhaftigkeit gerade zwischen Vokalen **sehr schwer zu erkennen**.

Der Lateral [l] – in vielen Sprachen übrigens zum Vokal [u] gewandelt oder als vokalähnlicher Approximant ausgesprochen – passt sich insbesondere in den hohen Formanten den umgebenden Vokalen an. Ohne auditive Kontrolle ist er sehr schwer rein visuell zu erkennen.



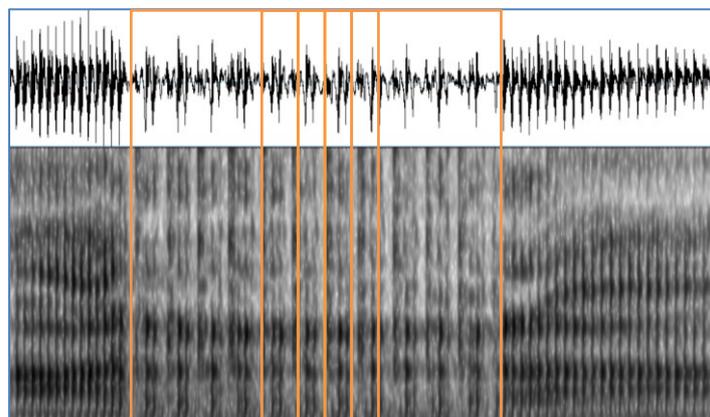
[l]

Ähnlich schwer zu erkennen ist das frikativisch ausgesprochene <r>: [ʁ] (streng genommen ein Frikativ und kein Vibrant). Auch hier sind die Schwärzungen leicht „heller“. Und auch hier ist oft auditive Kontrolle nötig. (Das [ʁ] hier wurde übrigens unüblich lange ausgesprochen, damit es besser sichtbar ist.)

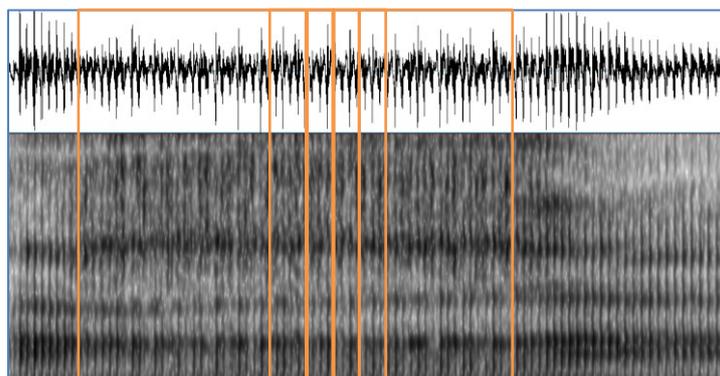


[ʁ]

Die beiden ‚richtigen‘ Vibranten [r] („Zungenspitzen-r“) und [R] („Zäpfchen-r“) sind wesentlich leichter zu erkennen. Das regelmäßige „Flattern“ der Zungenspitze (bei [r]) bzw. des Zäpfchens (bei [R]) ist klar im Spektrogramm abgebildet. Da das Zäpfchen kleiner ist und schneller flattert als die Zungenspitze, sind entsprechend mehr Schwingungen sichtbar.



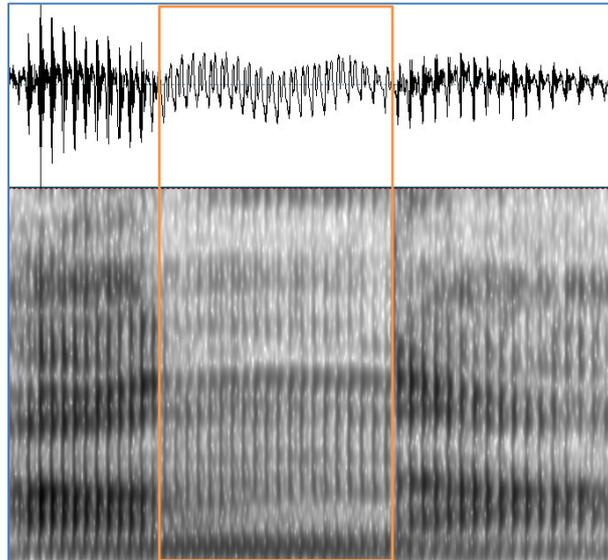
[r]



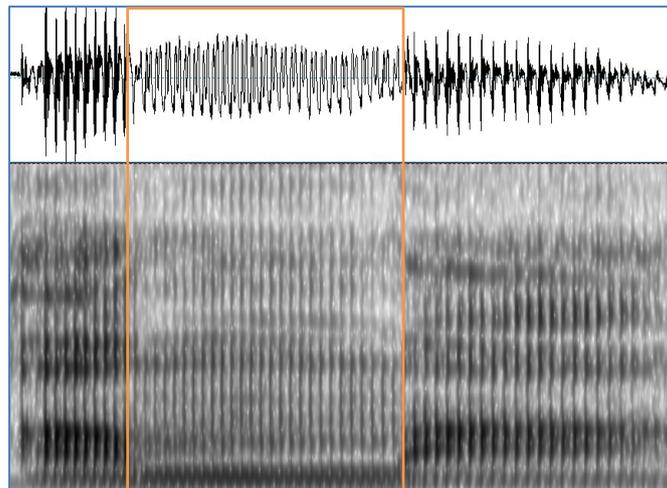
[R]

5. NASALE [N, M, ŋ]

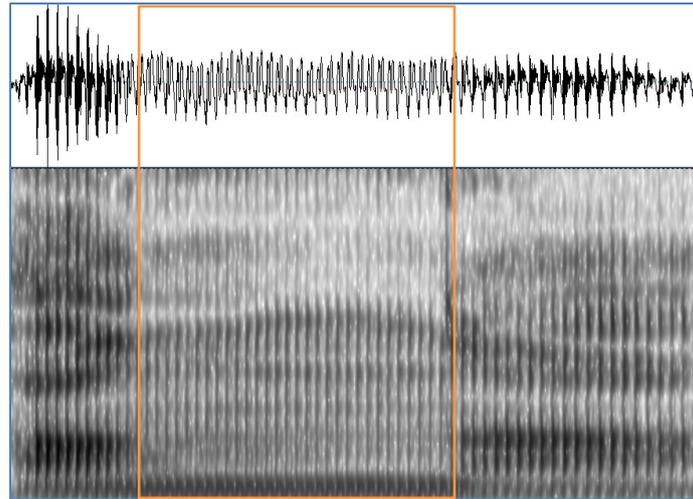
Die Nasale – Laute also, bei denen das Velum (=Gaumensegel) gesenkt ist und dadurch der Mundraum geschlossen, die Luft entweicht durch die Nase – zeichnen sich im Spektrogramm dadurch aus, dass die ‚voice bar‘ besonders stark ausgeprägt ist, während die mittleren und höheren Frequenzen kaum Energie aufweisen. Die unterschiedlichen Nasale lassen sich leider nur schwer unterscheiden (in der Regel durch Formanttransitionen bei den Vokalen), häufig muss geraten werden.



[n]



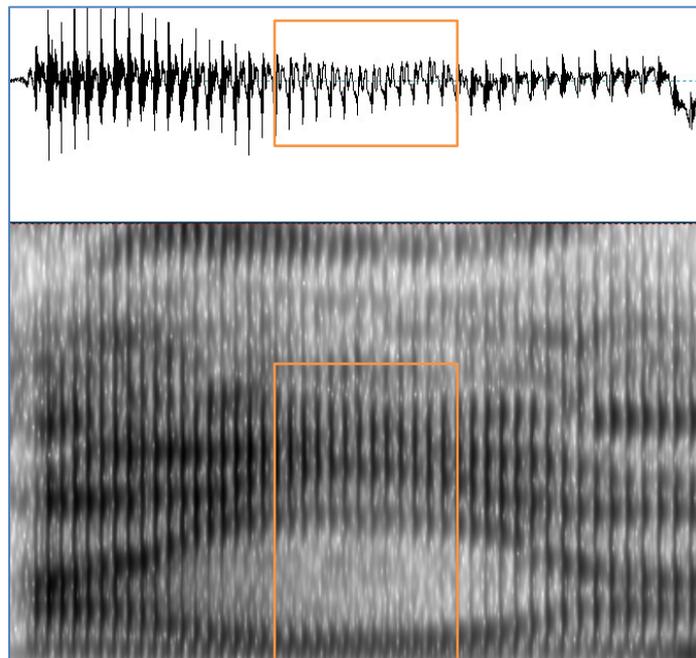
[m]



[ŋ]

6. APPROXIMANT [j]

Der einzige deutsche Approximant [j] verhält sich ähnlich – so auch die Definition eines Approximanten – wie ein Vokal (hier [i]). Dadurch ist er nur schwer von den Vokalen zu unterscheiden. Hinweise sind jedoch im **Oszillogramm** und in den **Formanten** zu finden: Beide sind blasser bzw. dünner, was durch die leichte Friktion des Approximanten (im Gegensatz zum Vokal) bedingt ist. Auch die Phonotaktik, also die Kombinationsmöglichkeiten der Laute in deutschen Silben und Wörtern) kann hier helfen, indem die möglichen Auftrittskontexte von [j] (in der Regel im Wort- bzw. Silbenanlaut) bei der visuellen Bestimmung mit herangezogen werden.



[j]

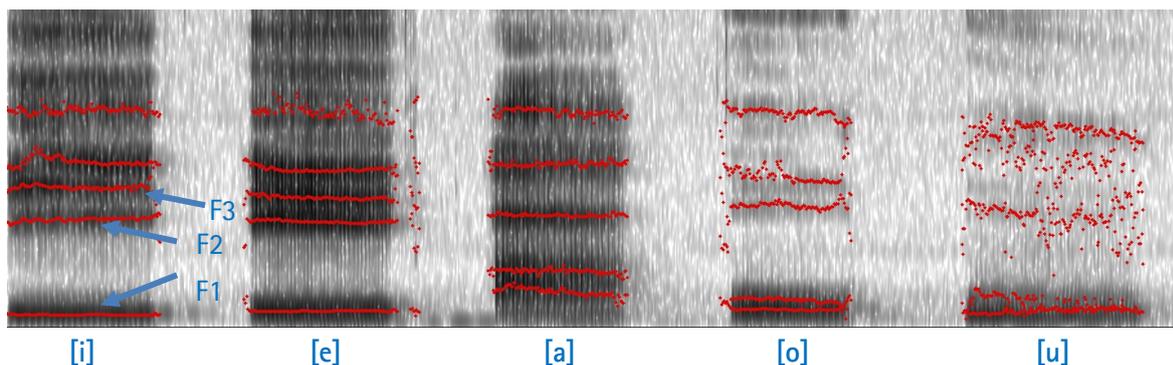
B) VOKALE UND DIPHTHONGE

1. VOKALE

Vokale unterscheiden sich von den Konsonanten (ausgenommen die Approximanten, die ähnlich aussehen) klar durch die so genannten ‚Formanten‘: Balken mit erhöhter Energiekonzentration, die anhand ihrer in der Regel **klaren Schwärzungen** leicht zu erkennen sind. Diese Frequenzbänder ergeben im Hinblick auf die Struktur der Schwärzungen häufig eine Art „Fischgrätenmuster“.

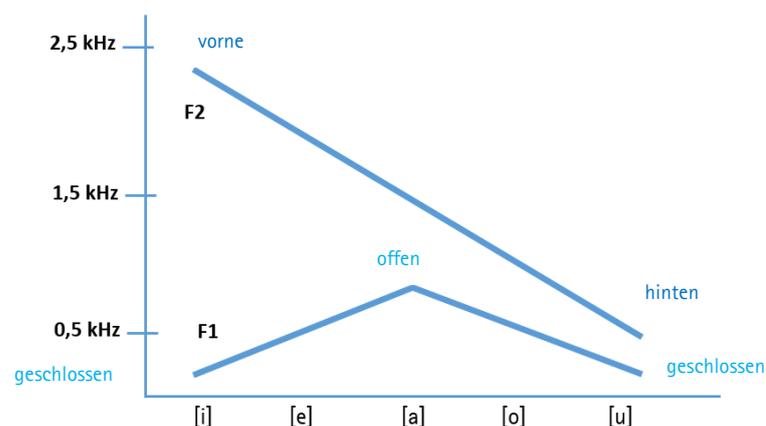
In PRAAT sind die Formanten anhand **kleiner roter Punkte**, die zusammen solche Balken darstellen, gut zu erkennen. Die wichtigsten dieser Formanten sind **F1** und **F2**:

- F1 korreliert mit dem **Öffnungsgrad des Mundes**: je offener der Mund, desto höher F1 (etwa bei [a]), je geschlossener, desto tiefer F1 (etwa bei [o])
- F2 korreliert mit der **horizontalen Zungenlage**: je weiter vorne ein Vokal artikuliert wird, desto höher F2 (etwa bei [i]), je weiter hinten, desto tiefer F2 (etwa [u])

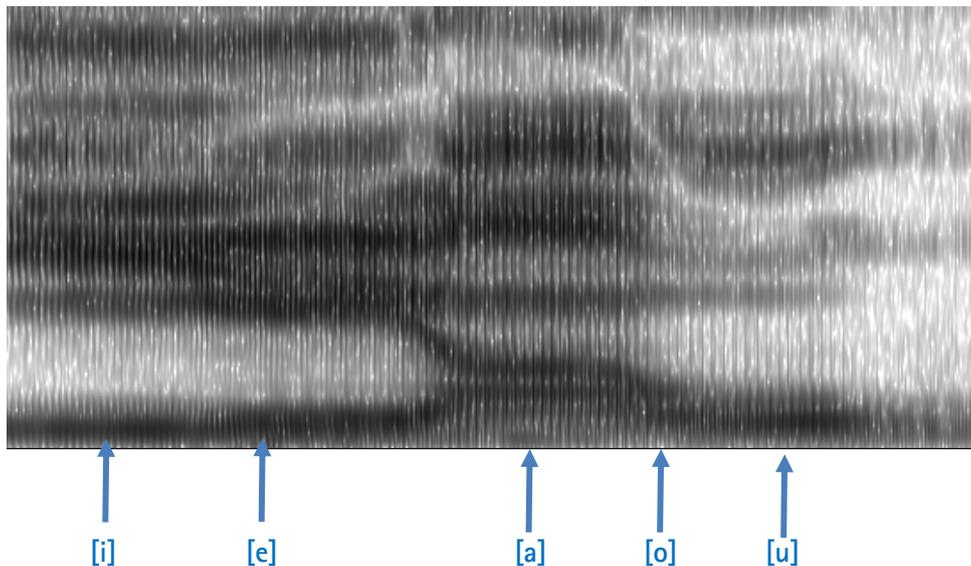


Gelegentlich sind die Balken eher wirr (wie in der Darstellung beim [u]). Aus Gründen der Koartikulation mit umgebenden Lauten sind sie zudem häufig nicht horizontal, sondern „in Bewegung“ („**Transitionen**“, siehe weiter unten bei den Diphthongen).

Schematisch lassen sich die Unterschiede zwischen **F1** und **F2** wie folgt darstellen, was zugleich – einmal abgespeichert – eine wichtige und schnell abrufbare Referenz ist, um die Vokale visuell zu bestimmen:



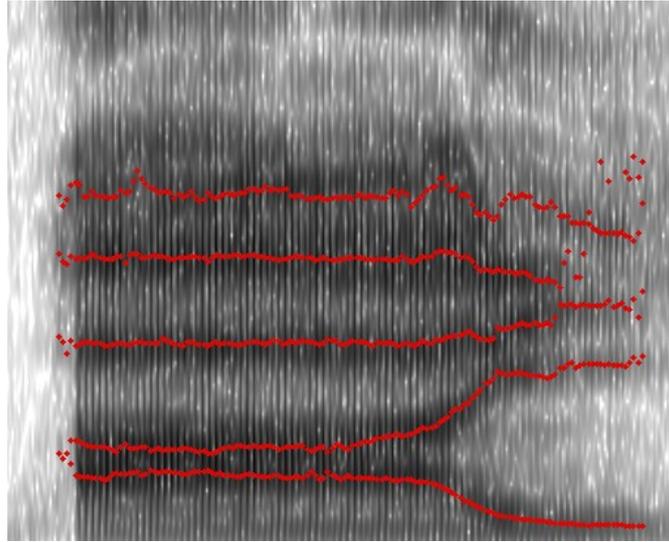
Diese Darstellung lässt sich auch direkt im Sonogramm erkennen, etwa in der folgenden Abfolge von [i – e – a – o – u]:



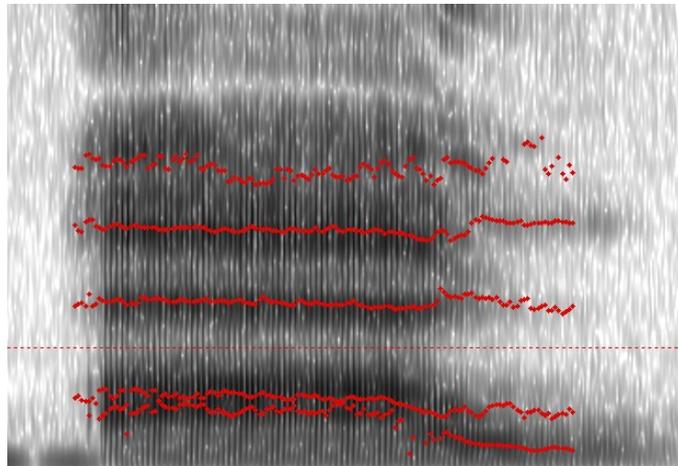
Somit lassen sich die einzelnen Vokale grundsätzlich anhand dieser beiden Parameter leicht unterscheiden. Dies gilt jedoch vor allem für die **Kardinalvokale** (sehr häufige Vokale, die sich am Rand des Vokaltrapezes befinden und daher in der Regel als Referenz für die restlichen Vokale in diesen Trapezen genommen werden). Bei einem komplexen Vokalsystem wie dem Deutschen ist es hingegen schwierig, die Qualitätsunterschiede rein visuell auf dem Spektrogramm auszumachen (etwa der Unterschied zwischen ungerundeten und gerundeten Gegenstücken wie [i] und [ɪ]). Je nach Sprecher sind sich auch die Kardinalvokale sehr ähnlich, etwa [i] und [e] im obigen Beispiel. Gerade wenn man das Vokalsystem der untersuchten Sprache kennt, ist die Formantlage jedoch ein zuverlässiger Indikator für die Qualität der einzelnen Vokale – insbesondere, wenn die Vokale jeweils **im Verhältnis zu den anderen Vokalen** analysiert werden (eine rein visuelle Verwechslung mit einem in dieser Sprache nicht vorkommenden Vokal ist kaum möglich).

2. DIPHTHONGE

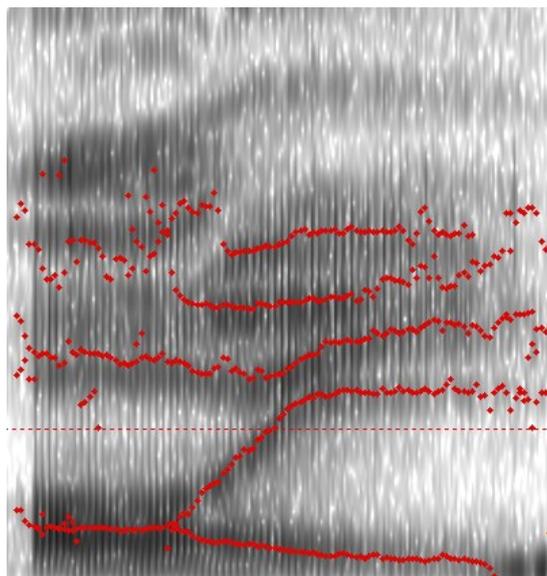
Diphthonge, die per Definition aus zwei Lauten bestehen, zeichnen sich dadurch aus, dass die Formanten sich „bewegen“, also von einem Vokal in den anderen übergleiten. Die **drei** im Deutschen vorkommenden Diphthonge [ai], [au] und [ɔɪ] sind in folgendem Sonogramm dargestellt. Insbesondere [ai] lässt sich aufgrund der klaren Formantbewegungen leicht erkennen (bei den beiden ersten Diphthongen ist das [a] länger gesprochen worden, beim dritten Diphthong ist das [ɔ] generell eher kürzer). Wo genau der eine Vokal aufhört und der andere beginnt, ist hingegen häufig schwer zu bestimmen.



[a ɪ]



[a ʊ]



[ɔ ɪ]

5) VOKALTRAPEZE

1. VORBEMERKUNGEN

- Bei Aufnahmen sollte die lautliche Umgebung für die dargestellten Vokale möglichst gleich sein (etwa fiktive Wörter wie *hatt, hutt, hitt*).
- Die beste Umgebung ist [h], weil hier der Mundraum am wenigsten verändert wird
- Umgebung, die vor allem zu vermeiden ist: Vor/Nach Liquiden und Nasalen > diese Laute haben starke **koartikulatorische Einflüsse** auf die Vokale und „verfärben“ diese
- am besten von jedem Vokal **mehrere Token**, optimal sogar verschiedene Wiederholungen desselben Wortes (die Werte werden dann gemittelt)
- **je mehr Vokal-Token, desto besser!**
- bei PRAAT werden die Formanten über „**Formant**“ > „**Show formants**“ als rote Punkte dargestellt, die sich leicht (automatisch oder manuell) auslesen lassen

2. FORMANTEN MESSEN

- **Messpunkte**, entweder
 - **genau in der Mitte des Vokals** (striktes Ausschließen der Koartikulation), entweder ein einzelner Formant oder ein Abschnitt > geht schnell
 - in **regelmäßigen Abständen** (z.B. immer 10ms) > braucht mehr Zeit und ist manchmal unsicher, weil Vokale unterschiedlich lang sind
 - in festgelegten, **proportionalen Abständen**, z.B. 20%, 50% und 80% des Vokals > beste Methode, weil für alle Vokale gleich, dadurch gut vergleichbar
- Bei den nun folgenden Beispielen wurde jeweils in der **Mitte des Vokals ein Abschnitt** markiert:
 - über die **F1-Taste** auf der Tastatur gibt PRAAT für diesen Bereich automatisch den gemittelten F1-Wert (erster Formant) aus
 - über die **F2-Taste** auf der Tastatur gibt PRAAT für diesen Bereich automatisch den gemittelten F2-Wert aus (zweiter Formant)
 - wichtig bei diesem Vorgehen ist, dass die Formanten sich im gewählten Abschnitt in etwa den gleichen Frequenzen befinden und nicht „wirr“ umherstreuen (was gelegentlich vorkommt) > hier ist **einige Erfahrung nötig**, um dies einwandfrei zu erkennen
 - im **Zweifelsfall: nur einen Messwert nehmen** (= einen jeweiligen roten Punkt für F1 und einen, gleich darüber!, für F2 wählen und die Frequenz notieren)

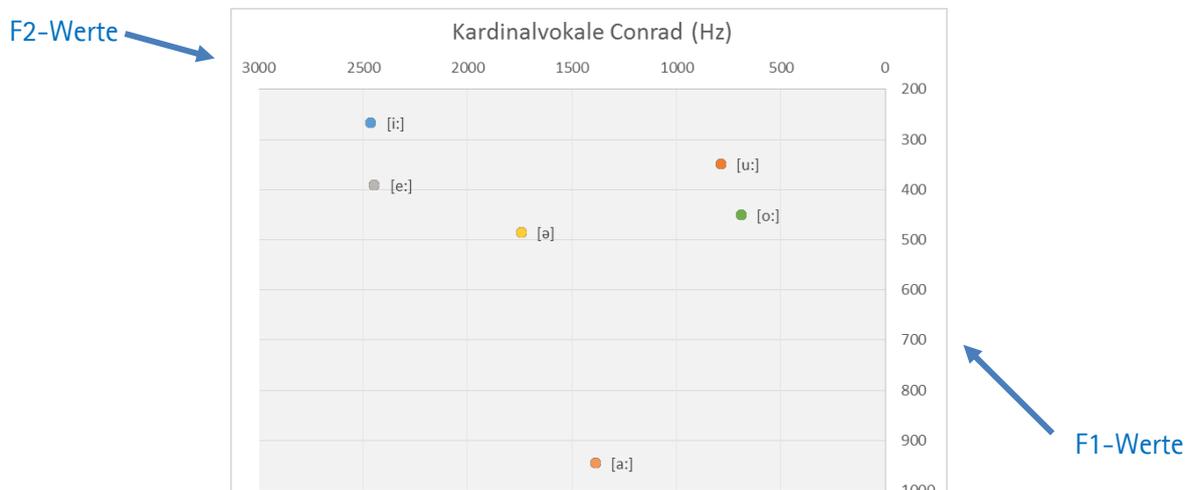
3. EXCEL-TABELLE ANLEGEN

Zu verzeichnen sind **Auftrittskontext** (hier: Wort), **Vokal** (hier Langvokale und Schwa) – am besten in IPA-Zeichen –, **F2** (korreliert mit horizontaler Zungenlage, also hinten-vorne) und **F1** (korreliert mit Öffnung des Mundes, also geschlossen-offen, bzw. vertikaler Zungenlage, also hoch-tief). Für die Verarbeitung in EXCEL ist es wichtig, dass F2 in der linken, F1 in der rechten Spalte steht. Nur so erscheinen die F2-Werte automatisch oben im Vokaltrapez, die F1-Werte rechts. Im anderen Fall müssen diese beiden Ebenen anhand mehrerer Mausklicks umgedreht werden – diese Arbeit kann man sich problemlos sparen.

Wort	Vokal	F2	F1
hiht	[i:]	2463	267
huht	[u:]	787	349
hoht	[o:]	690	451
haht	[a:]	1388	945
heht	[e:]	2446	391
heute	[ə]	1743	486

4. VOKALTRAPEZ ERSTELLEN

- 1) in EXCEL: Einfügen > PUNKTDIAGRAMM
- 2) Als „**Reihenname**“ bei der Auswahl der Werte das entsprechende Vokalsymbol eingeben (damit man sie später als Beschriftung einfügen kann)
- 3) Die **F2-Werte** kommen in die **x-Achse** und werden über dem Vokaltrapez dargestellt
 - Werte von **3000-0** (hierfür „Werte **in umgekehrter Reihenfolge** anklicken“)
 - Abstände von **500 Hz**
- 4) Die **F1-Werte** kommen in die **y-Achse** und werden in der Regel rechts vom Vokaltrapez dargestellt (hier gibt es aber auch Abweichungen)
 - Wert von **200-1000** (hierfür „Werte **in umgekehrter Reihenfolge** anklicken“, kleinere Werte kommen in der Regel nicht vor)
 - Abstände von **100 Hz**
- 5) Falls die **Beschriftungen** der einzelnen Punkte nicht automatisch erscheinen (was häufig der Fall ist) > manuell einfügen (manchmal geht das über „**Wert aus Zellen angeben**“, dann kann die Zeile mit den IPA Symbolen auch angegeben werden)
- 6) Wahlweise kann man die Punkte auch **mit Linien verbinden**



5. NORMALISIERUNG DER WERTE

In der Regel werden die Werte nicht in Hertz dargestellt, sondern in einer "normalisierten" Einheit wie BARK. Hintergrund hiervon ist, dass sich z.B. Männer, Frauen und Kinder aufgrund des **unterschiedlich großen Ansatzrohres** in ihren Hertz-Werten zum Teil stark unterscheiden. Unterschiede zwischen diesen SprecherInnengruppen könnten daher allein auf **anatomische Differenzen** zurückzuführen sein. Werden Daten in andere Einheiten umgerechnet, werden diese Unterschiede **nivelliert** und die Werte somit **vergleichbar** gemacht.

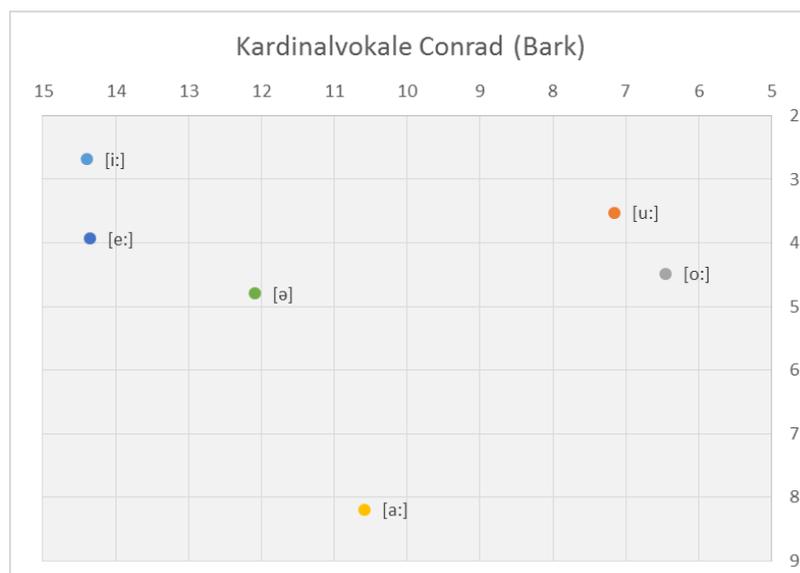
Die häufigste solcher Einheiten ist **BARK**. Die von Hertz in Bark umgewandelten Werte bewegen sich auf einer Skala von in der Regel **2-9** (auf der y-Achse) bzw. **5-15** (auf der x-Achse). Perzeptionspsychologisch werden Vokale, die sich um mehr als 1 BARK unterscheiden, als unterschiedliche Vokale wahrgenommen. Vokale hingegen, die sich um **weniger als 1 BARK unterscheiden**, werden als **gleiche Vokalqualitäten wahrgenommen**.

Die Umrechnung erfolgt nach folgender Formel, wobei f den jeweiligen Hertz-Wert darstellt:

$$z = [26,81 / (1 + 1960 / f)] - 0,53$$

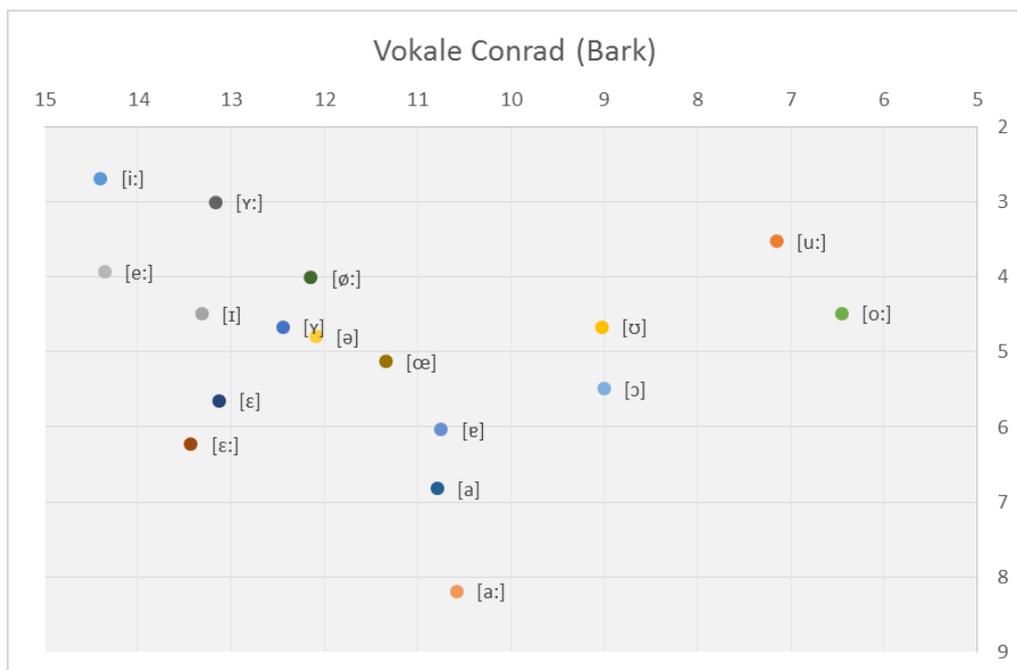
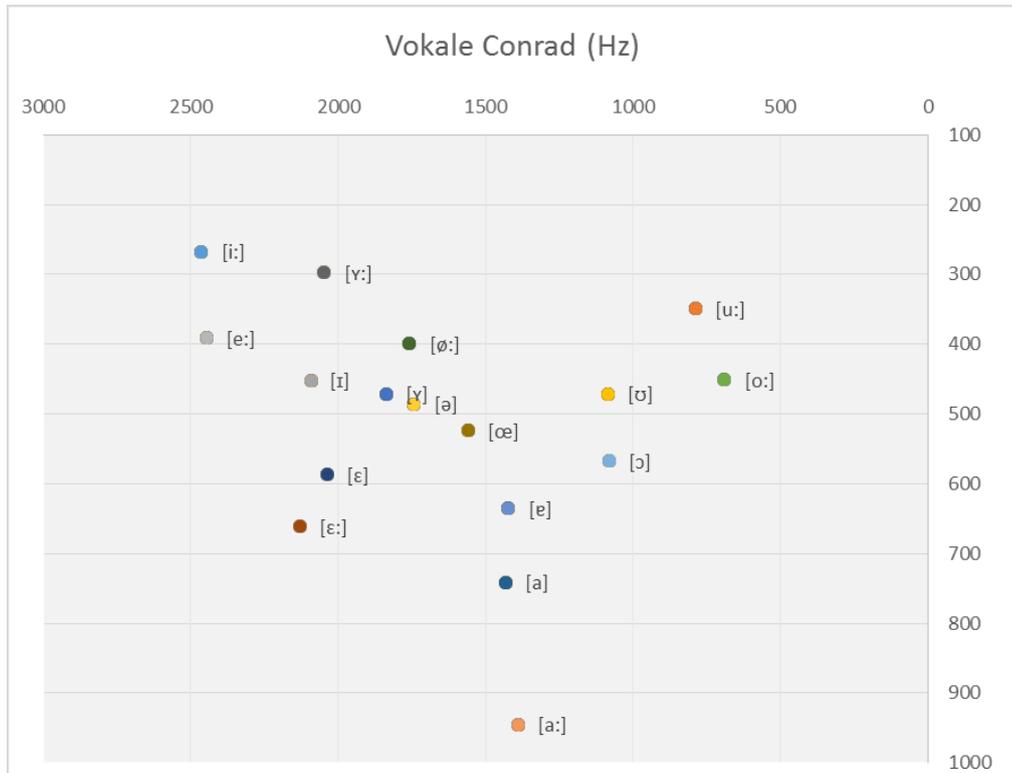
Was nach komplizierter Mathematik aussieht, lässt sich in EXCEL ganz entspannt ausrechnen. Ist die Formel einmal eingegeben, lässt sich der jeweilige Hertz-Wert problemlos einsetzen.

Das normalisierte Vokaltrapez unterscheidet sich in unserem Beispiel nur leicht von dem nicht-normalisierten, da nur wenige Daten von nur einem Sprecher gezeigt werden. Bei mehreren (vor allem bei vielen!) verschiedenen Vokalen und SprecherInnen, die miteinander verglichen werden, ist eine **Normalisierung dringend zu empfehlen!**



6. KOMPLEXE VOKALTRAPEZE

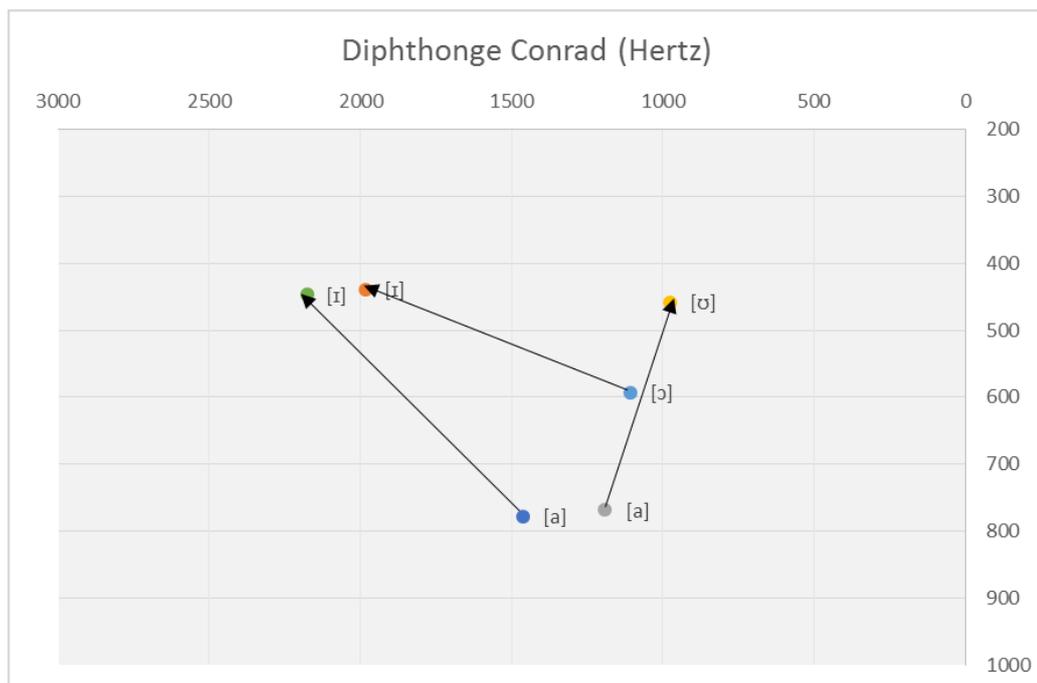
Auch komplexe Vokaltrapeze, wie sie insbesondere für das Deutsche nötig sind, lassen sich auf die gleiche Weise darstellen. Werden die Kardinalvokale durch die restlichen deutschen Vokale ergänzt, ergeben sich folgende Darstellungen (erneut in HERTZ und in BARK):



7. DARSTELLUNG VON DIPHTHONGEN

Bei den Diphthongen ist es nötig, **mindestens zwei Messwerte für F1 und zwei für F2** zu haben: Einen von jedem Laut des Doppellautes.

Es empfiehlt sich jedoch, mehrere Messungen (etwa in festgelegten Abständen) vorzunehmen, da sich dann die Bewegung der Zunge von einem zum nächsten Vokal klarer darstellen lässt. Wenn nur der Anfangs- und Endpunkt bzw. die „Grundqualität“ der beteiligten Vokale interessiert, reicht ein Messwert pro Vokal aus, wie im folgenden Beispiel (*heute, Haut, -heit*). Die Pfeile wurden **manuell eingezeichnet**.



8. LITERATUR

DI PAOLO, Marianna / YAEGER-DROR, Malcah (Hgg.): *Sociophonetics. A student's guide*. London/New York: Peter Lang. **Kapitel 8** ('Analysing vowels') und **Kapitel 9** ('More on vowels: plotting and normalization').

Auf [youtube](#) gibt es auch kurze Videos zum Erstellen von Vokaltrapezen mit EXCEL.

Und schließlich: Es gibt auch fertige **EXCEL-Templates** (über die gängigen Suchmaschinen zu finden) zum Fertigen von Vokaltrapezen. Einfach die F1- und F2-Werte sowie die dazugehörigen Wörter einfügen (diese müssen im angeführten Beispiel relativ aufwändig umformatiert werden), der Rest übernimmt das Programm automatisch. Geht natürlich viel schneller – dafür hat man aber auch weniger den Durchblick für die einzelnen Schritte. Nicht zufällig steht dies also erst am Ende dieser Beschreibung ...

6) VOICE ONSET TIME (VOT)

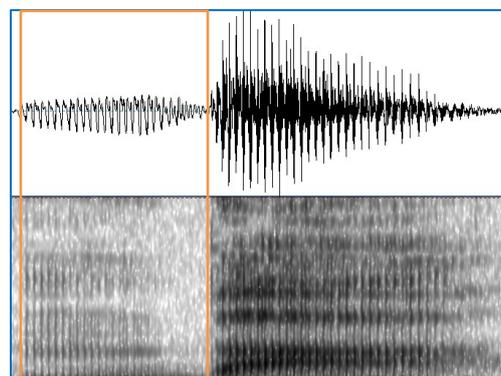
Ein viel untersuchtes und in der Regel auch sicher feststellbares Unterscheidungsmerkmal der Plosive ist die so genannte **VOICE ONSET TIME (VOT)** – also die Zeitspanne ab der Verschlusslösung bis zur Schwingung der Stimmlippen (=Phonation) beim folgenden Vokal. Die VOT wird in drei Kategorien eingeteilt:

PRE-VOICED: Phonation tritt bereits **während des Verschlusses** auf und bleibt bis zum folgenden Vokal erhalten (im Deutschen kein phonologisches Merkmal, häufig aber z.B. in asiatischen Sprachen wie dem Vietnamesischen, was sich in der lateinischen Umschrift wie in *Nguyen* äußert)

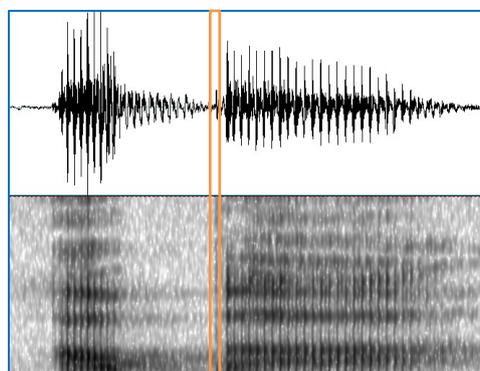
UNBEHAUCHT/NICHT-ASPIRIERT: Phonation tritt unmittelbar **nach der Verschlusslösung** auf (dies betrifft im Deutschen alle stimmhaften Plosive, also /b, d, g/)

BEHAUCHT/ASPIRIERT: Die Phonation tritt **nach einer frikativischen „Aspirationsphase“** erst **mit oder kurz vor dem Vokal** auf (typisch für die deutschen stimmlosen Plosive /p, t, k/)

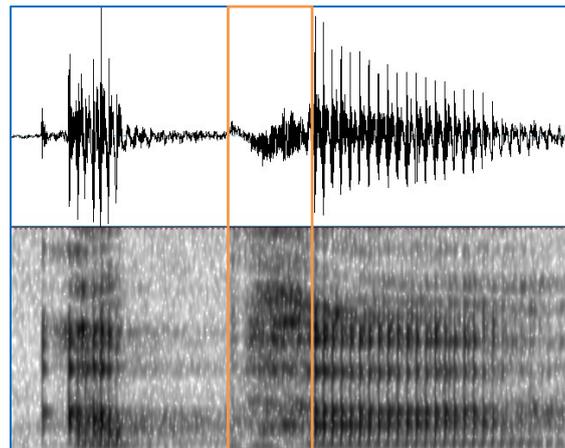
Die folgenden Darstellungen zeigen jeweils ein Beispiel für die drei Kategorien (Beispielwörter: *mba*, *aba*, *apa*):



[mba] (pre-voiced)



[aba] (unbehaucht)



[ap^ha]

Typischerweise bewegen sich die **durchschnittlichen VOT-Werte** der drei Kategorien wie folgt (bei pre-voiced werden negative VOT-Werte angenommen). Diese Werte variieren jedoch von Sprache zu Sprache (und von SprecherIn zu SprecherIn):

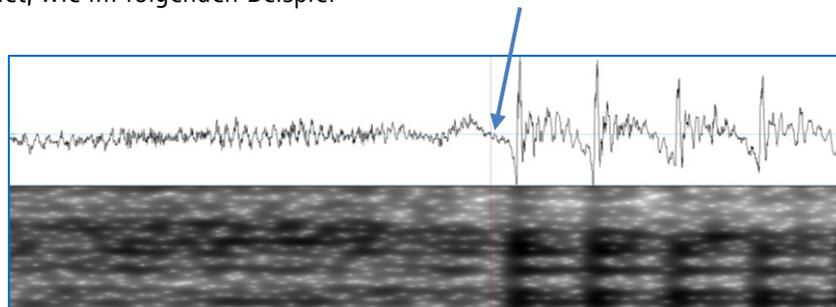
PRE-VOICED:	-125 bis -75 ms
NICHT-BEHAUCHT/ASPIRIERT:	0-25 ms
BEHAUCHT/ASPIRIERT:	ab 25 ms

Beeinflusst wird die VOT unter anderem durch

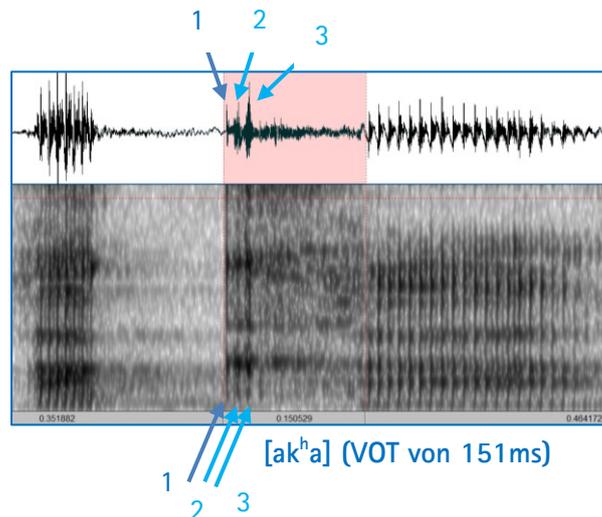
- den Artikulationsort (höher je weiter hinten artikuliert)
 - die Kontaktfläche der Zunge mit dem Mundraum
 - die Betonung: stimmlose Plosive in betonten Silben haben eine höhere VOT als solche in unbetonten Silben
- } daher $p < t < k$

Umstritten ist auch die **Messmethode**: Wann genau beginnt die Phonation beim anschließenden Vokal? Wann genau beginnt die Verschlusslösung? Als Faustregel können folgende Regeln gelten:

- die Darstellung im Oszillogramm ist wichtiger als diejenige im Spektrogramm
- die Messung erfolgt immer am so genannten „Nulldurchgang“ – wo also die Kurve die y-Achse schneidet, wie im folgenden Beispiel



- bei **multiplen Verschlusslösungen** (siehe Beispiel unten, dort 3 Stück) empfehle ich stets den ersten zu nehmen (auch hier gibt es aber unterschiedliche Deutungen)



- und, ganz generell: für alle Messungen die gleichen Parameter anwenden!

Da die Methoden in unterschiedlichen Forschungen oft unterschiedlich sind, ist eine Vergleichbarkeit häufig schwer möglich. Unbedingt ist die angewandte Methode bei der Lektüre anderer Forschung zu prüfen und bei Interpretationen mit zu berücksichtigen! Mit zahlreichen unklaren Fällen ist bei einer Analyse von VOT immer zu rechnen.

LITERATUR:

DI PAOLO, Marianna / YAEGER-DROR, Malcah (Hgg.): *Sociophonetics. A student's guide*. London/New York: Peter Lang. Kapitel 6, S. 62-68.

7) ANLEITUNGEN, ONLINE-MATERIALIEN ETC.

ONLINE-TUTORIALS:

- <http://www.fon.hum.uva.nl/praat/manualsByOthers.html>
- besonders zu empfehlen (auf Deutsch): <http://praatpfanne.lingphon.net>
- vieles lässt sich auch ergoogeln ...

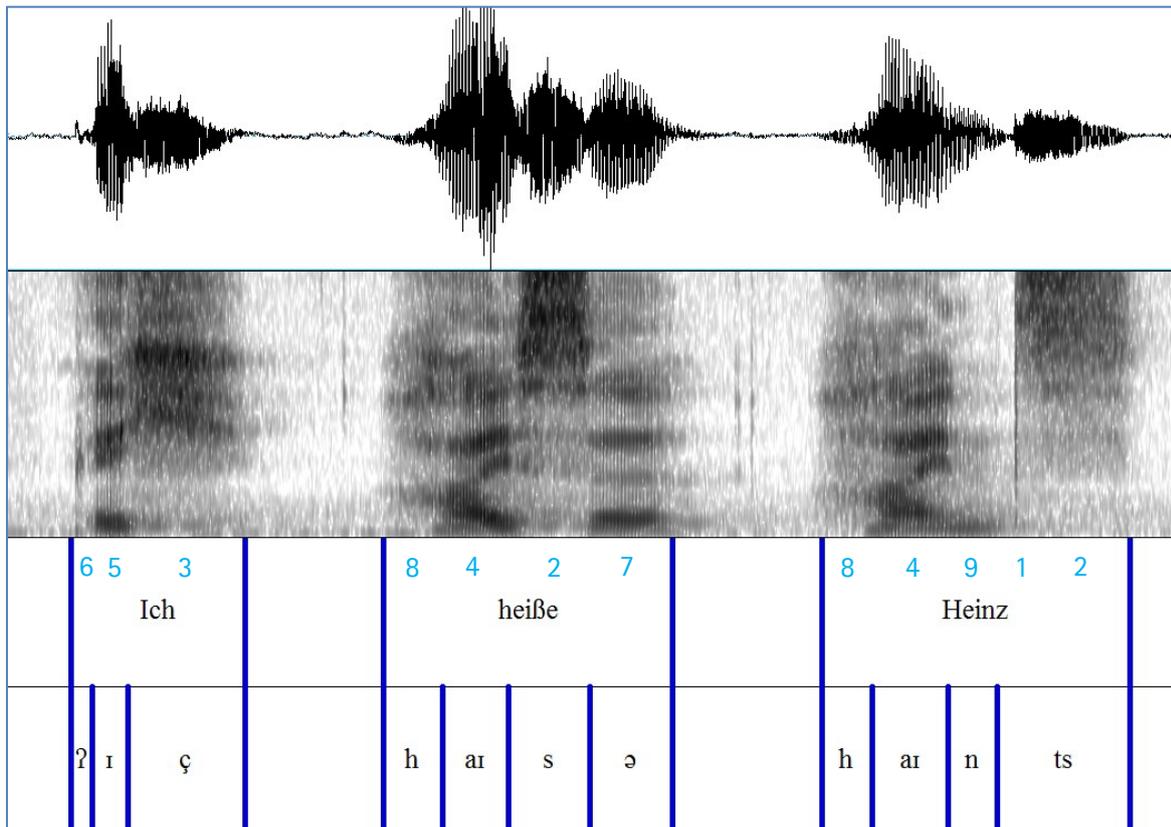
(SOZIO-)PHONETISCHE BÜCHER MIT VIELEN DETAILS ZU EINZELLAUTEN:

- POMPINO-MARSCHALL, Bernd (2009): *Einführung in die Phonetik*. 3., durchgesehene Auflage. Berlin/New York: de Gruyter.
- DI PAOLO, Marianna / YAEGER-DROR, Malcah (Hgg.): *Sociophonetics. A student's guide*. London/New York: Peter Lang.
- THOMAS, Erik. R. (2011): *Sociophonetics. An introduction*. Houndmills: Palgrave.

HILFE-FUNKTION bei PRAAT:

- oben rechts „Help“ > „Praat Intro“
- leider m.E. unnötig unübersichtlich und kompliziert, dafür aber komplett

8) AUFLÖSUNG DER „STUMMEN“ ANALYSE



Kommentar (Analyseschritte in chronologischer Reihenfolge):

Am leichtesten zu erkennen sind die (1) stimmlosen Plosive [p, t, k] durch die Verschlussphase und die Verschlusslösung, wie das [t] in *Heinz* (in der Kombination [ts] = Affrikate; hier keine Aspiration, weil der Plosiv in den Frikativ [s] übergeht). Ebenfalls in der Regel klar zu identifizieren ist (2) das [s] durch die starken Schwärzungen im oberen Frequenzbereich, wie in *heiße* und *Heinz* (zweiter Teil der Affrikate [ts]). Auch der palatale Frikativ [ç] in *ich* (3) ist aufgrund seiner typischen Musterung unproblematisch. Vokalisch ist der Diphthong [ai] (4) durch die klare Bewegung der Formanten „aus der Mitte nach oben“ gut zu erkennen (*heiße*, *Heinz*).

Mit diesem Grundgerüst lassen sich die noch übrigen Laute leichter identifizieren bzw. raten. Die Formanten im ersten Vokal (5) (bei *ich*) deuten im Vergleich zum Diphthong auf ein [i] hin. Vor dem anlautenden Vokal ist ein Knacklaut/glottaler Plosiv [ʔ] zu erkennen (6). Den Schwa-Vokal [ə] bei *heiße* (7) erkennt man an den regelmäßigen Formantlagen. [h] ist grundsätzlich eher schwer zu erkennen (in *heiße* und *Heinz*), kann jetzt aber leicht kombiniert werden (8). Das [n] ist hier eher kurz, weist aber die typisch dicke voice bar und kaum Schwärzungen in den mittleren und höheren Frequenzbereichen auf (9).