

16. Stevens K. The quantal nature of speech: evidence from articulatory-acoustic data / Edward E. David Jr., Peter B. Denes // Human Communication: A Unified View. – New York : McGraw Hill, 1972. – S. 51–66.
  17. Stevens K. On the quantal nature of speech / K. Stevens // Journal of Phonetics. – Band 17. – 1989. – S. 3–46.
  18. Stevens K. Acoustic Phonetics / K. Stevens. – Cambridge Mass.: The MIT Press, 1999. – 607 S.
  19. Tabain M. Articulation and acoustics of /i/ in preboundary position in French / M. Tabain, P. Perrier // Journal of Phonetics. – Band 33. – 2005. – S. 77–100.
  20. Wiesinger P. Die Standardaussprache in Österreich / P. Wiesinger // Deutsches Aussprachewörterbuch. – Berlin; New York : de Gruyter, 2009. – S. 229–258.
  21. Wood S. A radiographic analysis of constriction location for vowels / S. Wood // Journal of Phonetics. – Band 7. – 1979. – S. 25–43.
  22. Wood S. X-Ray and Model Studies of Vowel Articulation / S. Wood // Lund Working Papers. – Band 23. – 1982.
- Wood S. The Acoustical Significance of Tongue, Lip, and Larynx Maneuvers in Rounded Palatal Vowels / S. Wood // Journal of the Acoustical Society of America. – Band 80. – 1986. – S. 391–401.

\*\*\*\*\*

***Pompino-Marschall, Bernd  
Humboldt-Universität zu Berlin***

### **WAS ZEIGT UNS DAS SONAGRAMM?**

*У статті детально представлено можливості використання програмного забезпечення Praat для проведення акустичного аналізу мовлення. Надано практичні поради щодо сегментації й анотації сонаграм на власному ПК, проаналізовано основні фізичні форми й параметри звуків, деталізовано акустичну артикуляцію та специфічні акустичні зразки звукових сигналів.*

**Ключові слова:** акустичний аналіз, сонаграма, форманти, анотація, транскрипція.

*В статье детально представлены возможности использования программного обеспечения Praat для проведения акустического анализа звучащей речи. Даются практические рекомендации относительно сегментации и аннотации сонаграмм с помощью компьютера, проанализированы основные физические формы и параметры звуков, детализированы акустическая артикуляция и специфические акустические образцы звуковых сигналов.*

**Ключевые слова:** акустический анализ, сонаграмма, форманты, аннотация, транскрипция.

*The article presents a detailed description of Praat software capacities for the acoustic analysis of speech. Given are practical pieces of advice for segmenting and summarizing sonograms using one's own PC, analyzed are basic physical forms and parameters of sounds, specified are acoustic articulation and particular acoustic samples of sound signals.*

**Key words:** acoustic analysis, sonogram, formants, summary, transcription.

## 1. Prolog: Akustische Sprachanalysen am eigenen PC

Für die Bearbeitung signalphonetischer Fragen am eigenen audiofähigen PC/Laptop ist das unter <http://www.praat.org> als Freeware erhältliche Programmpaket PRAAT von Paul Boersma und David Weenink (1992-2016) zu empfehlen.

Dieses Programmpaket, das inzwischen auch von vielen Forschungseinrichtungen standardmäßig (mit)genutzt wird, ermöglicht akustische Analysen, phonetische Segmentation und Etikettierung sowie Signalmanipulationen, wie dies vor ein paar Jahrzehnten nur bei extrem teurer Spezialsoftware gegeben war.

Da dieses überaus empfehlenswerte Programm für normale Windows-Programme gewohnten Nutzer zugegebenermaßen einige ungewohnte Eigenschaften aufweist, sei an dieser Stelle eine sehr knappe Arbeitsanleitung für die grundlegenden Funktionen gegeben.<sup>1</sup> Ich beschränke mich hierbei im Wesentlichen auf die akustische Signalerfassung und die Möglichkeiten des Abhörens sowie der Segmentation und Etikettierung anhand des Osziollogramms und des Sonagramms und schließlich der Grundfrequenz- und Spektralanalyse.

### *Nach dem Start von PRAAT*

Nach dem Programmstart erscheinen auf dem Bildschirm standardmäßig zwei Fenster: *Praat objects* (links) und *Praat picture* (rechts). Das Objekt-Fenster ist dabei das für die weitere Arbeit maßgebliche. Mit seinen Menü-Funktionen können Sie z.B. neue Signale aufnehmen (*New*), auf der Festplatte gespeicherte Audiodateien unterschiedlicher Formate (u.a. längere \*.wav Dateien) für die Weiterbearbeitung öffnen (*Open*), Signal- und Analysedateien in unterschiedlichen Formaten auf Festplatte abspeichern (*Save*)<sup>2</sup> sowie unter dem Menüpunkt *Help* auf verschiedene Tutori als zugreifen.

### *Einlesen einer Sprachsignaldatei*

Für die weitere Verarbeitung in PRAAT können auf Platte abgespeicherte Audiodateien unterschiedlichsten Formats<sup>3</sup> mit *Open – Read from file* (bzw. *Read – Open long sound file*) eingelesen oder aber über das Menü *New (New – Record mono/stereo Sound)* neu aufgenommen werden.<sup>4</sup>

Der PRAAT *SoundRecorder* (ein eigenes Fenster) bietet die Möglichkeit unterschiedlicher Abtastraten bei der Aufnahme – standardmäßig die halbe Abtastrate (22,05 kHz) des Audio CD Standards –, die für die akustische Sprachanalyse völlig ausreichend ist. (Nach der Aufnahme mit dem PRAAT SoundRecorder muss die Datei – ggf. umbenannt – mit dem Menüpunkt *Save to list* in die Liste der PRAAT-Objekte übernommen werden.) Wenn keine spektrale Frikativanalyse geplant ist, reicht auch eine Abtastrate von 11,025 Hz aus bzw. steigert sogar die Realibilität der Grundfrequenz- und Formantanalysen.

---

<sup>1</sup> Der Leser sei zudem auf das sehr hilfreiche englischsprachige Tutorial von Pascal van Lieshout verwiesen, das ebenfalls von der PRAAT-Homepage aus zugänglich ist. Auch die Help-Funktion von PRAAT bietet Tutorials, ist ansonsten aber eher ein Handbuch für den in digitaler Signalverarbeitung versierteren Nutzer. Für deutschsprachige Leser ist v.a. die "Praatpfanne" von Jörg Meyer (<http://praatpfanne.lingphon.net>) zu empfehlen.

<sup>2</sup> N.B.: Die im Objekt-Fenster während der Arbeit mit PRAAT gelisteten und nach Windows-Standard auswählbaren Objekte befinden sich grundsätzlich im Computer-Memory und müssen vor dem Schließen des Programms zur Weiterverwendung ggf. erst noch auf Platte abgespeichert werden (PRAAT warnt Sie hier vorsichtshalber).

<sup>3</sup> Es empfiehlt sich, generell das standardgemäße Windows \*.wav Format zu verwenden. Komprimierende Formate wie MPEG sind N.B. für akustisch-phonetische Analysen ungeeignet.

<sup>4</sup> Hierzu sind vorab windowsgemäß die Eingangskanäle und deren Aussteuerung zu spezifizieren.

TIPP: Digitalisieren Sie Ihre Audiodaten am besten mittels des Windows-Zubehörs (oder anderer Programme) als längere \*.wav Dateien (in PRAAT: *long sound file*) mit einer Abtastfrequenz von 22,05 kHz und sichern Sie diese auf CD. Lesen Sie diese Daten sodann als *long sound files* in PRAAT ein. Nutzen Sie sodann PRAAT, um die für Ihre Weiterarbeit ggf. downgesampelten Teildateien zu erzeugen.

#### *Weiterverarbeitung, Visualisierung und Abhören einer Audiodatei*

Wenn Sie eine Audiodatei aus der Liste der PRAAT-Objekte auswählen (durch Anklicken), verändert sich das Menü des Objekt-Fensters.<sup>5</sup> Neben den nun aktiven Menüpunkten unterhalb des Listenfeldes links, mit denen Sie Dateien umbenennen (*Rename*), aus der Liste löschen (*Remove*), kopieren (*Copy*) bzw. sich Informationen über diese geben lassen können (*Info, Inspect*), erscheinen auf der rechten Seite die objektspezifischen Menüpunkte. Um z.B. die Abtastrate eines Signals zu verändern, wählen Sie aus dem unteren rechten Menüabschnitt *Convert Resample*, woraufhin sich ein Dialogfenster öffnet, in dem Sie die neue Abtastrate eingeben können.<sup>6</sup> Nach der Abarbeitung des Befehls erscheint dann in der Objekt-Liste ein weiteres Objekt mit dem Namenszusatz, gefolgt vom Abtastfrequenzwert (z.B. kommt zu *Sound bass* das Objekt *Sound bass\_10000*).

Eine ausgewählte lange Audiodatei können Sie über den Menüpunkt *View*, eine normale ausgewählte Audiodatei über den Menüpunkt *Edit* in einem Fenster, das Ihnen das Oszillogramm und Sonagramm zeigt, weiter verarbeiten (vgl. Abb. 1).

Das Editierfenster beinhaltet oben eine Menüleiste und darunter ein Oszillogramm- und Sonagrammfenster, in denen auch zusätzliche Analyseparameter wie Grundfrequenz- und Intensitätsverlauf, Formanten und Periodenmarkierungen angezeigt werden können.<sup>7</sup> Mittels Mausclick können Sie im Signalfenster einen Cursor setzen und mittels einer horizontalen Bewegung der gedrückt gehaltenen Maus innerhalb der Signalfenster können Sie – ähnlich wie in Textverarbeitungssystemen – Abschnitte auswählen, die dann rosa unterlegt erscheinen. In dieser Einstellung haben sie nun die Möglichkeit, einzelne Signalabschnitte mittels der Abspieltasten (vgl. Abb. 1) gezielt abzuhören: Die Taste oberhalb der Signalfenster gibt die Auswahl wieder, ebenso die entsprechende in der ersten Reihe unterhalb der Signalfenster, die zudem Tasten für das vorangehende und nachfolgende Signal innerhalb des Fensters aufweist. Mit der mittleren Taste hören Sie das gesamte im Fenster sichtbare Signal und mit der untersten das gesamte im Memory verfügbare Signal.<sup>8</sup>

Unterhalb der Abspieltasten finden sich links die Zoomtasten, mit denen sich das dargestellte Signal zeitlich dehnen (*in*), stauchen (*out*) und auf Gesamtdarstellung des Memory-Inhalts einstellen (*all*) lässt. Die Taste *sel* dehnt den ausgewählten Signalebereich auf die gesamte Fensterbreite. Rechts neben den Zoomtasten erscheint dabei die Scrollbar, mit der Sie das Signal im Fenster verschieben können.

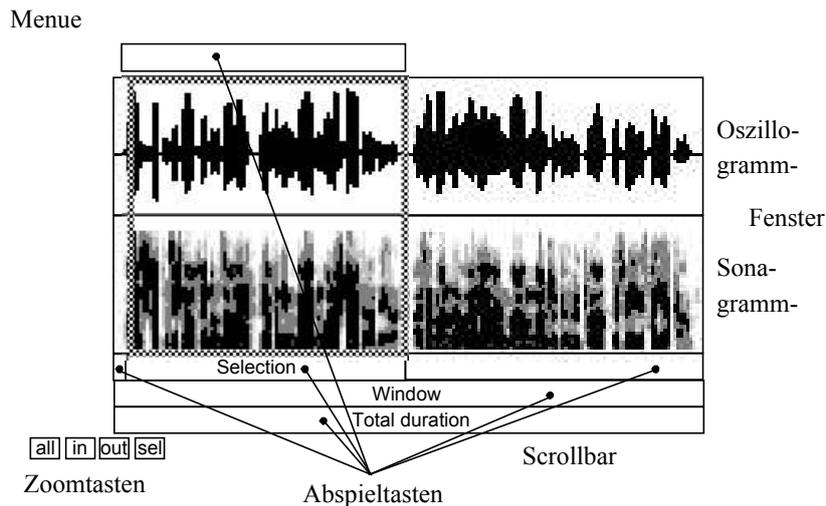
---

<sup>5</sup> Das Objekt-Fenster zeigt immer ein dem ausgewählten Objekt entsprechendes Menü.

<sup>6</sup> Meist sind in den PRAAT Dialogfenstern häufig verwendete Default-Werte (bzw. Ihre eigene letzte Einstellung während der Arbeit mit PRAAT) bereits vorgegeben. Die auch hier immer vorhandene Helpfunktion führt zu den jeweils relevanten Handbuchausschnitten.

<sup>7</sup> Diese sind über den Menüpunkt *View* aktivierbar.

<sup>8</sup> Sie können die Audiowiedergabe dabei jederzeit durch Drücken der Escape-Taste abbrechen.



**Abb. 1: Das PRAAT Sound Edit-Fenster (vgl. Text zu Details)**

Für die Weiterverarbeitung langer Audiodateien empfiehlt sich eine weitere Möglichkeit der Signalauswahl mittels des Menüpunktes *Select*: Hier können Sie den Cursor sowie den Beginn/das Ende des Signalausschnitts auf definierte Zeitpunkte setzen bzw. um definierte Zeitintervalle verschieben. Mit den Menübefehlen *File – Extract selected sound ...* bzw. *File – Save selected sound as ...* können Sie zur Weiterbearbeitung Signalausschnitte ins Memory laden bzw. auf Platte abspeichern.

TIPP: Erzeugen Sie zur weiteren Bearbeitung aus Ihren langen Audiodateien auf diese Weise überlappende (evtl. downgesampelte) Audiodateien gleicher Länge für die anschließende Segmentation und Etikettierung.<sup>9</sup>

#### *Segmentieren und Etikettieren von Sprachsignaldateien*

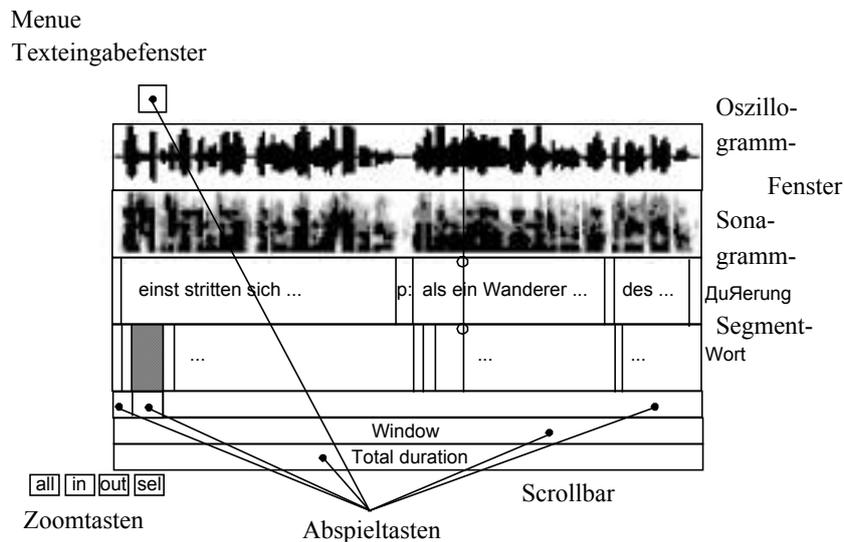
Am häufigsten werden Sie PRAAT wohl als Transkriptionswerkzeug verwenden. Hierzu müssen Sie einer ausgewählten Audiodatei eine Annotationdatei zuordnen. Dazu verwenden Sie den Befehl *Annotate – To TextGrid* im oberen Abschnitt des rechten Menüfeldes, der wiederum ein Dialogfenster erscheinen lässt, in dem Sie die durch Leerzeichen voneinander getrennten Namen der Segmentationsebenen (*Tier names*) eingeben können und in einer zweiten Zeile bestimmen, welche dieser Ebenen nur Zeitpunktinformationen enthalten sollen (*Point tiers*).<sup>10</sup> Wählen Sie anschließend im PRAAT Objekt-Fenster (mit gleichzeitig gedrückter Control-Taste) die Audiodatei und die gleichnamige TextGrid-Datei. Unter dem Menüpunkt *Edit* erhalten Sie nun ein dem oben beschriebenen Sound-Editierfenster äquivalentes Segmentationsfenster mit zusätzlichen Fenstern für die angegebenen Segmentierungsebenen (vgl. Abb. 2).

Die Cursorpositionierung (Mausklick in den Signalfenstern) mit den oben beschriebenen Abhörfunktionen führt nun zur Möglichkeit der Segmentation: In den unterhalb der Signalfenster liegenden Segmentfenstern erscheint an der Cursorposition

<sup>9</sup> Die zeitliche Überlappung sollte so gewählt werden, dass zumindest eine gemeinsame Segmentgrenze in den aufeinander folgenden Dateien vorhanden ist, damit eine fortlaufende zeitliche Zuordnung zum originalen langen Audiofile möglich bleibt. N.B.: Bei der Abspeicherung eines mittels *File – Extract selection (preserve times)* erhaltenen Audioobjekts auf Platte geht die Zeitinformation verloren.

<sup>10</sup> Die nicht unter *Point tiers* genannten Ebenen speichern die Segmentationsinformationen als Anfangs- und Endgrenzen. Hier ist die Default-Einstellung (*Mary John bell* bzw. *bell*) natürlich Unsinn. Eine von Ihnen gewünschte Annotation könnte aber z.B. folgendermaßen aussehen: *Wort Laut Tonposition* bzw. *Tonposition*.

jeweils oben ein hellblauer Kreis (vgl. Abb. 2, rechte Bildhälfte), der durch (präzises) Anklicken in eine (verschiebbare) Segmentgrenze im TextGrid wird.<sup>11</sup> Haben Sie entsprechende Grenzen gesetzt, so können Sie die Segmente mittels des oben unter der Menüleiste befindlichen Texteingabefensters annotieren. Der TextGrid enthält sodann die Zeitinformation (Anfangs- und Endgrenze bzw. Zeitmarke) und Annotation zu ihrem Audiosignal. Sie können abgespeicherte TextGrids immer wieder mit dem Audiosignal gemeinsam öffnen und haben Ihr Segmentationsergebnis vor Augen. Im Segmentationsfenster haben Sie auch eine weitere Abhörmöglichkeit:



**Abb. 2: Das PRAAT-Segmentationsfenster (vgl. Text zu Details)**

Wenn Sie in den Segmentfenstern ein bestimmtes Segment anklicken, wird dieses ausgewählt (vgl. Abb. 2, linke Bildhälfte) und Sie können es durch Drücken der Tabulatortaste abspielen.<sup>12</sup>

Auf die Darstellung weiterer Funktionen von PRAAT sei hier vorerst verzichtet. Wir wollen uns nun vielmehr der Frage zuwenden, was wir in den Darstellungen des akustischen Sprachsignals – v.a. im sogenannten Sonagramm – sehen können und wie uns diese Informationen bei der Aufgabe der Segmentierung des Sprachsignals helfen können.

## 2. Schall: Grundformen und Analyseparameter<sup>13</sup>

Schall sind von uns auditiv wahrnehmbare minimale Luftdruckschwankungen. Um für uns wahrnehmbar zu sein, müssen diese Schwingungen in einer gewissen Geschwindigkeit ablaufen, und zwar zwischen 20 bis 20000 mal in der Sekunde, d.h. mit einer Frequenz zwischen ca. 20 Hz [Hertz] und 20 kHz. Bezüglich ihrer Amplitude sind diese wahrnehmbaren Luftdruckschwankungen als äußerst gering, aber einen großen Intensitätsbereich umspannend zu kennzeichnen. So liegt die absolute Hörschwelle bei den vom Menschen am besten wahrnehmbaren Frequenzen um 3500 Hz bei einer Druckschwankung von ca.  $10^{-4}$   $\mu\text{b}$  [Mikrobar] (einem Zehntausendstel

<sup>11</sup> Das TextGrid ist über *File – Save TextGrid as text file* als ASCII-File abzuspeichern.

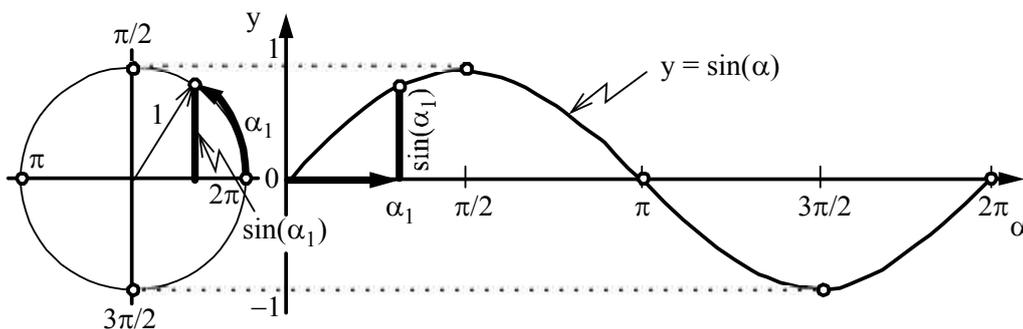
<sup>12</sup> Für die ohrenphonetische Beurteilung/Transkription sollten Sie jedoch darauf achten, jeweils längere Abschnitte abzuhören, da scharfe Segmentgrenzen zu auditiven Artefakten führen können.

<sup>13</sup> Für eine ausführlichere Darstellung vgl. Pompino-Marschall (2009).

von einem Millionstel des uns umgebenden Luftdrucks von ca.  $1 \text{ b}^{14}$ , d.h.  $0,0000000001 \text{ b}$ ), die Schmerzgrenze hingegen bei einem Wert von ca.  $10^3 \text{ } \mu\text{b}$  (tausend Millionstel des atmosphärischen Drucks, d.h.  $0,001 \text{ b}$ ). Letzteres ist der praktische Grund, weshalb der Schalldruck im logarithmischen Maß Dezibel [dB] ausgedrückt wird: Eine Veränderung des Schalldruckpegels um  $\pm 6 \text{ dB}$  entspricht dabei einer Verdoppelung/Halbierung des Schalldrucks,  $20 \text{ dB}$  einer Änderung des Schalldrucks um den Faktor 10.

Schall tritt uns in unterschiedlichen Ausprägungsformen entgegen: Als stimmhafter, periodischer Schall in Form von Tönen und Klängen, als aperiodischer Schall in Form von Geräuschen und Transienten (Clicks und Knallen).

Als Ausgangspunkt für die physikalische Analyse von Schall dient die "klassische" Form einer Schwingung, aus der sich alle beliebigen Schwingungsformen ableiten lassen, die Pendel- oder Sinoidalschwingung. Es ist dies die Bewegung, die ein Pendel, einmal aus seiner Ruhelage gebracht (bei Fehlen von dämpfenden Reibungskräften) unaufhörlich um seine Ruhelage herum vollführen würde: Es bewegt sich mit maximaler Geschwindigkeit durch den Punkt seiner Ruhelage, bremst bis zum Punkt der Maximalauslenkung auf Nullgeschwindigkeit ab, um von da an bis zum neuerlichen Durchschreiten der Ruhelage in entgegengesetzter Richtung zu beschleunigen und bis zum gegenüberliegenden Wendepunkt wiederum kontinuierlich abzubremesen u.s.f. Dieser absolut kontinuierliche Bewegungsprozess lässt sich anschaulich als sog. Kreisschwingung beschreiben, wie dies in Abb. 3 dargestellt ist:



**Abb. 3: Die Sinusschwingung als Kreisschwingung**

Der zeitabhängige Amplitudenwert dieser Schwingung ergibt sich durch den Abstand eines kontinuierlich entgegengesetzt zum Uhrzeigersinn auf dem Kreisumfang wandernden Punktes gegenüber der x-Achse. Bei einem Kreisradius von 1 (d.h. beim Einheitskreis) entspricht dies der bei 0 mit dem Wert 0 startenden und in Abhängigkeit vom Winkel  $\alpha$  (in Grad, d.h. Bruchteilen von 360 bzw. im Bogenmaß als Bruchteilen von  $2\pi$ , d.h. jeweils eines Kreisumfangs) zwischen den Werten  $\pm 1$  schwankenden Sinusfunktion, welche nach einer Kreisumrundung wiederum den Wert 0 erreicht. Die Geschwindigkeit, mit der um den Kreis geschritten wird, bestimmt dabei die Frequenz:

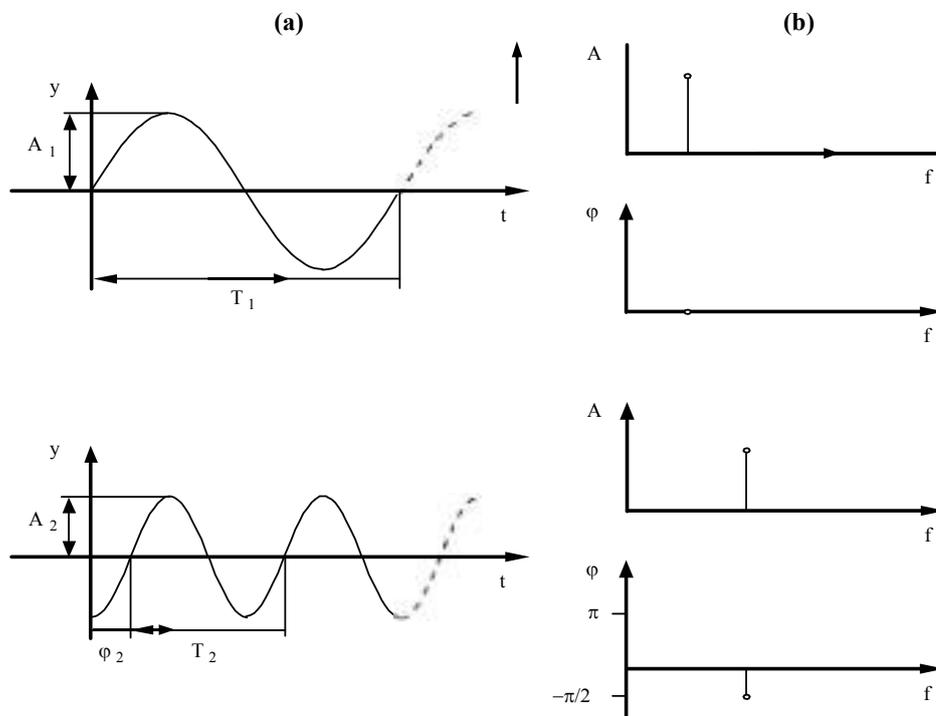
$$y(t) = A \cdot \sin(2\pi \cdot f \cdot t + \varphi)$$

<sup>14</sup> Der Einfachheit halber, d.h. wg. des Bezugs auf den atmosphärischen Druck, wurde hier die alte Einheit bar verwendet; die Darstellung in der heute gültigen Einheit Pa [Pascal] bedeutet demgegenüber lediglich eine Veränderung des Stellenwerts um 5;  $1 \text{ Pa} = 10^{-5} \text{ b}$ .

Zur Kennzeichnung einer beliebigen Sinoidalschwingung genügen somit die drei Parameterwerte

- Amplitude (A) als Multiplikator der selbst ja nur zwischen  $\pm 1$  variierenden sin-Funktion,
- Frequenz (t) als multiplikatives Element im Argument der sin-Funktion, die bei dem Teilargument ( $2\pi \cdot t[\text{sec}]$ ) innerhalb einer Sekunde einmal den Kreisumfang umschreibt (*Periode*), d.h. eine *Frequenz* von 1 Hz aufweist und
- Phase ( $\varphi$ ) als additives Element des Arguments, das angibt, um wieviel der Startpunkt der Schwingung – bezogen auf einen Schwingungszyklus – gegenüber dem Sinus mit Startpunkt 0 verschoben ist (so ist z.B. die cos-Funktion mit Startpunkt 1 gegenüber dem Sinus um  $90^\circ$  bzw.  $\pi/2$  verschoben).

Diese Zusammenhänge sind – gemeinsam mit der Darstellung von Sinoidalschwingungen im Zeit- und Frequenzbereich – in der folgenden Abb. 4 verdeutlicht:

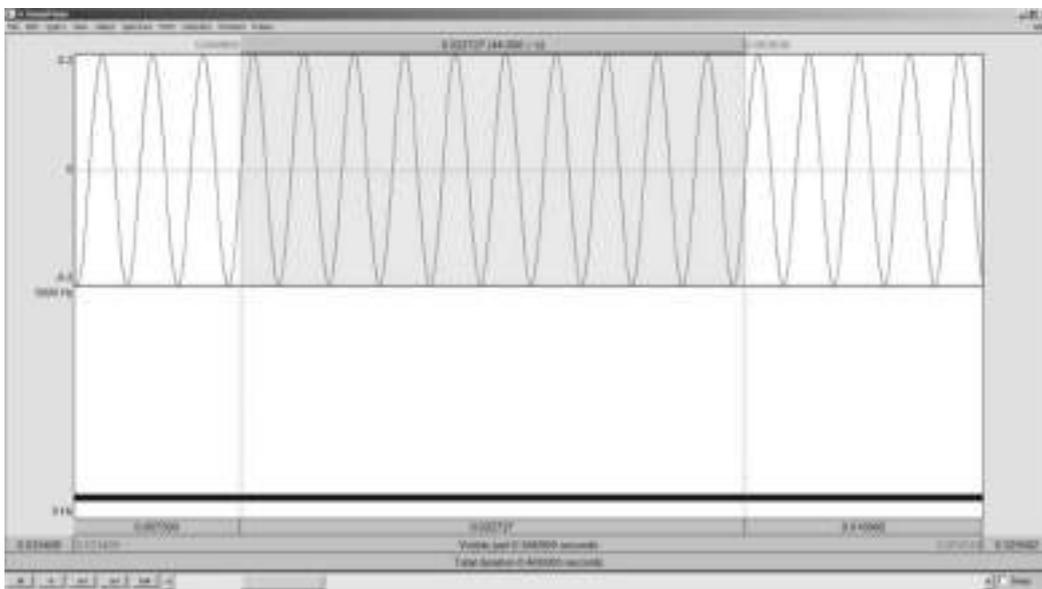


**Abb. 4: Die Kenngrößen Amplitude (A), Periodendauer (T) [bzw. Frequenz ( $f = 1/T$ )] und Phase ( $\varphi$ ) der Sinoidalschwingung in der Darstellung (a) des Zeit- und (b) Frequenzbereichs (vgl. Text)**

Unter (a) ist in der bekannten Weise der Zeitverlauf (das Oszillogramm) zweier Sinoidalschwingungen dargestellt (als Amplitudenveränderung über die Zeit), wobei die obere im gleichen Zeitraum (durchgezogener Schwingungsverlauf) einen, die untere zwei Schwingungszyklen durchläuft. Zudem beginnt die obere Schwingung positiv ansteigend bei 0, während die untere zum Zeitpunkt 0 ihre negative Maximalamplitude aufweist, d.h. einen Viertelzyklus vorher ( $-\pi/2$ ) bzw. einen Dreiviertelzyklus später ( $3\pi/2$ ) beginnt. Unter (b) sind dieselben Schwingungen im Frequenzbereich, d.h. bezüglich ihres Spektrums (vgl. u.) dargestellt (die Koordinatenwerte stehen nun einerseits für die Frequenz, andererseits für die Amplitude bzw. die Phase): die obere Schwingung hat ein Amplitudenspektrum von der Größe  $A_1$  an einer einzigen

bestimmten Frequenz mit dem dazugehörigen Phasenspektrum von 0, während die untere Schwingung in Entsprechung als Amplitude  $A_2$  bei der doppelten Frequenz der oberen Schwingung sowie einer Phase von  $-\pi/2$  bei dieser Frequenz darstellbar ist.

Auch in PRAAT können wir uns die Sinusschwingung nochmals verdeutlichen. In Abb. 5 ist mit *View & Edit* das Audiosignal dargestellt, das wir uns z.B. mit dem Menübefehl *New* → *Sound* → *Create Sound as pure tone ...* und den dazugehörigen Standardeinstellungen erzeugt haben: der Kammerton  $a^1$  (440 Hz). In der oberen Bildhälfte ist ein Ausschnitt von 50 ms als Oszillogramm der Schwingung zu sehen mit einer schattierten Auswahl von 10 Perioden à 22,727 ms (die 44/s ergäbe), in der unteren Bildhälfte sehen wir – als Entsprechung im Frequenzbereich – eine waagerechte Schwärzung auf Höhe 440 Hz im Schmalbandsonagramm (Fensterbreite 30 ms; Erläuterung s.u.).



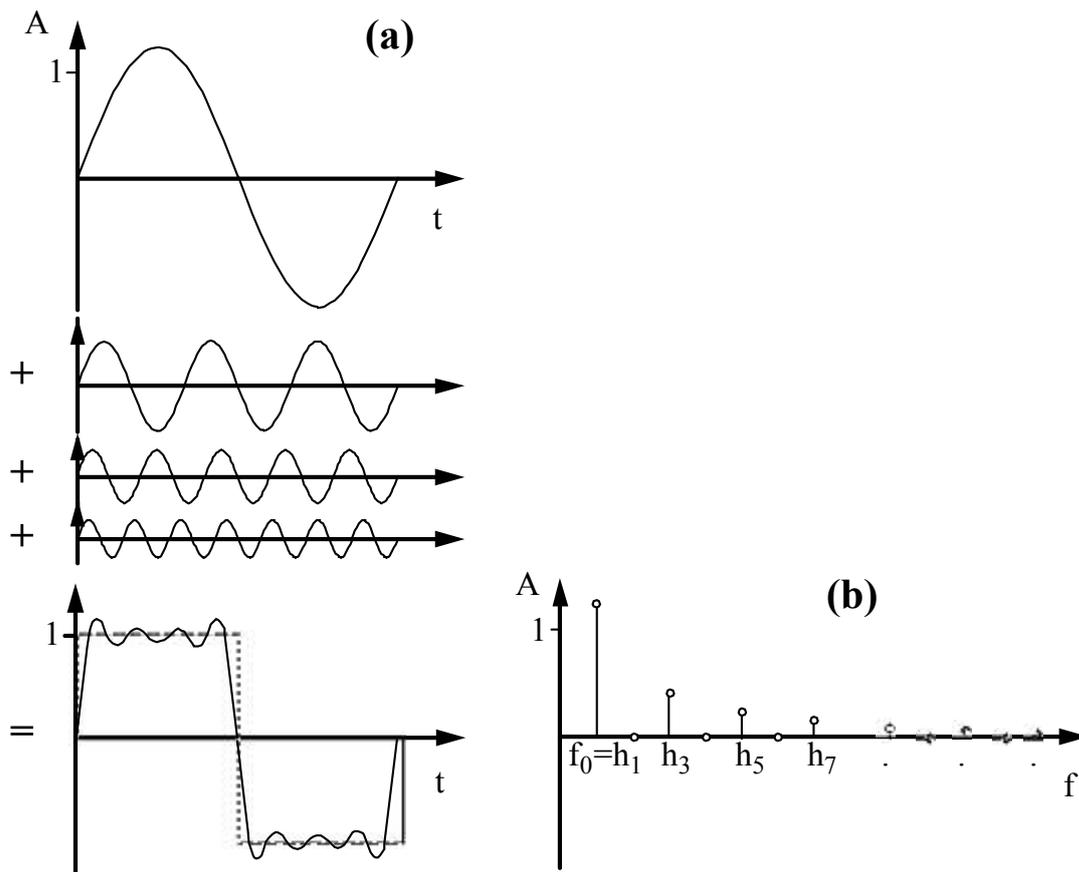
**Abb.5: Oszillogramm (oben) und Schmalbandsonagramm (unten) des Kammertons  $a^1$  (440 Hz) im PRAAT Editierfenster**

Eingangs des Kapitels wurde gesagt, dass die Sinusschwingung die Grundlage für die Analyse jedweder Schwingung bildet. Wie dies funktioniert, sei im Folgenden in zwei Schritten erläutert: an Klängen einerseits und nichtperiodischen Schwingungen andererseits.

Als Klänge (im engeren Sinn) bezeichnen wir alle die Schwingungen, die in Form eines Linienspektrums als Addition einzelner Sinuskomponenten zu analysieren sind, wobei diese Einzelkomponenten in einem harmonischen Verhältnis zueinander stehen (d.h. ein ganzzahliges Vielfaches einer gemeinsamen Grundfrequenz  $[f_0]$  darstellen):

$$y(t) = \sum_{n=1}^{\infty} A_n \cdot \sin(2\pi \cdot f_n \cdot t + \varphi_n)$$

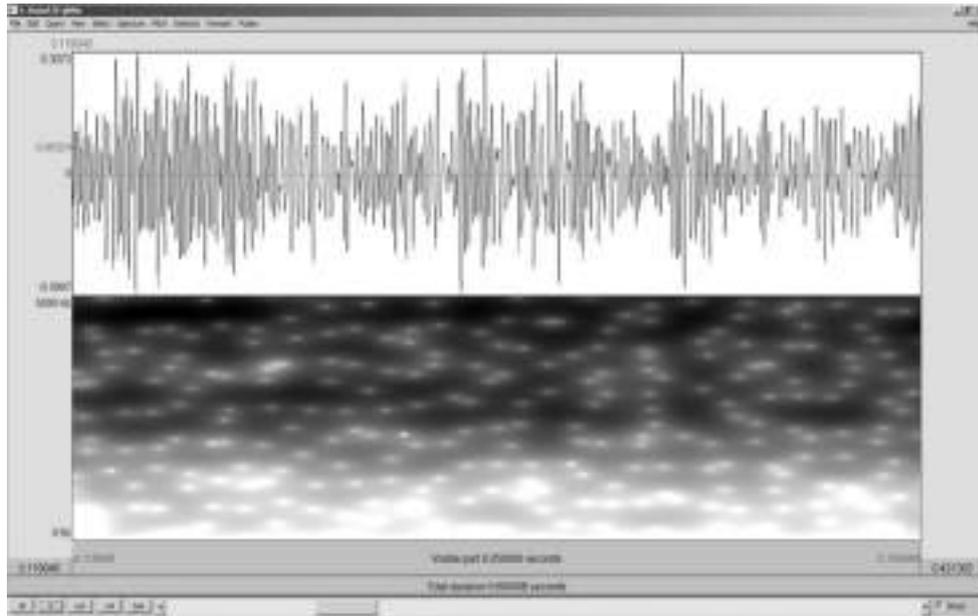
In Abb. 6 ist dies anhand einer Rechteckschwingung und ihrer teilweisen Annäherung durch die Addition einzelner Sinuskomponenten verdeutlicht.



**Abb. 6: Annäherung einer Rechteckschwingung durch die additive Überlagerung (Fourier-Synthese) harmonischer Sinusschwingungen in der (a) Zeitbereichs- und (b) Frequenzbereichsdarstellung (unter (b) sind einige der gegenüber der in (a) gepunktet eingezeichneten Rechteckschwingung fehlenden spektralen Komponenten ebenfalls angedeutet)**

Um die in der Abbildung 6 unten links gepunktet eingezeichnete rechteckige Schwingungsform anzunähern, bedarf es der Addition der jeweils ungeradzahligten Harmonischen der Grundfrequenz (d.h. der Sinuskomponente mit derselben Periodizität wie die der zu analysierenden Schwingung plus je einer Schwingung der drei-, fünf-, siebenfachen usw. Frequenz) mit bei aufsteigender Frequenz exponential abnehmender Amplitude und einer Phasenverschiebung von 0 (in Abb. 6 sind unter (a) die einzelnen Sinuskomponenten sowie ihre Summe im Zeitbereich dargestellt, unter (b) das resultierende Amplitudenspektrum).

Stellen wir uns nun eine nichtperiodische Schwingung – ein Geräusch – vor. “Nichtperiodisch” bedeutet gleichermaßen eine unendlich lange Periodendauer, die “Grund”-Frequenz dieser Schwingung wäre also unendlich niedrig. Bei einem nichtperiodischen Signal würden wir es so – parallel zu den Abb. 3 und 5 oben – im Frequenzbereich mit einem “Linien”-Spektrum mit unendlich eng beinaderliegenden Spektrallinien, einem sog. “kontinuierlichen” Spektrum zu tun haben. Dies sei vorausgreifend an einer PRAAT-Darstellung eines stimmlosen Frikativs verdeutlicht:



**Abb. 7: PRAAT-Darstellung eines Ausschnitts von 50 ms des Frikativs [ʃ] (aus šar)**

3. Akustische Artikulation: Das Zustandekommen der spezifischen akustischen Muster des Sprachschalls

Bei der natürlichen Sprachproduktion wird das Zusammenspiel von artikulatorischen und aerodynamischen Vorgängen in akustische Prozesse transformiert.<sup>15</sup> Hierbei geht es grundsätzlich um zwei zu unterscheidende Phänomenbereiche: Auf der einen Seite steht der wiederum durch zwei unterscheidbare Prozesse charakterisierbare Vorgang der *Rohschallerzeugung*, auf der anderen Seite der darauf folgende Vorgang der *Rohschallmodifikation/-modulation* durch die geometrieabhängige Filterfunktion des an die Quelle des Rohschalls anschließenden Teils des Ansatzrohres.

*Rohschallerzeugung: Phonation*

Die Art und Weise der Stimmlippenschwingung ist grundsätzlich von zwei unterschiedlichen Faktoren abhängig: den aerodynamischen Kräften der transglottalen Luftströme (wie sie durch den subglottalen Druck und die geometrische Form der glottalen Öffnung bedingt sind) einerseits und den von der Dicke, Spannung und dem Grad der Adduktion der Stimmlippen abhängigen myoelastischen Kräften andererseits. Das für die akustische Sprachproduktion wichtigste Moment besteht dabei in der periodischen Wiederkehr der durch die myoelastischen und die Bernoulli-Kräfte bewirkten Bildung abrupter Stimmlippenverschlüsse: Während bei geöffneter Glottis mehr oder weniger Luft aus der Lunge ausströmt, kommt es bei der ‚Abschneidung‘ dieses transglottalen Luftstroms durch den glottalen Verschluss direkt oberhalb der Glottis zu einer impulsartigen Störung des Luftdrucks in negativer Richtung, einer sog. Kavitätenbildung. Diese lokale Störung des Luftdrucks ihrerseits breitet sich mit Schallgeschwindigkeit in der oberhalb der Glottis gelegenen Luftsäule des Ansatzrohres aus (vgl. u.). Mit dem Stimmlippenverschluss kommt es also zu einem im Ansatzrohr reflektierten negativen Luftdruckimpuls, wohingegen die Luftdruckverhältnisse während der Öffnungsphase hauptsächlich durch den transglottal fließenden Luftstrom bedingt sind.

<sup>15</sup> Für eine ausführlichere Darstellung vgl. Pompino-Marschall (2009).

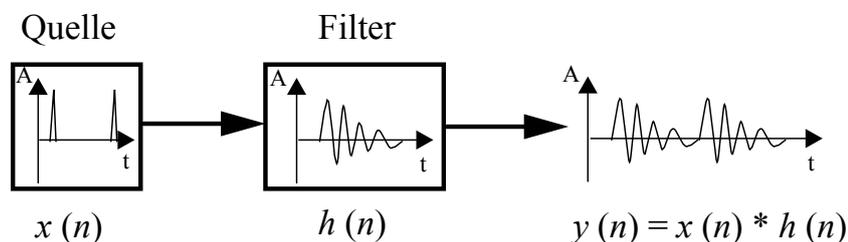
### *Rohschallerzeugung: Friktion*

Die zweite Form der Rohschallerzeugung beruht auf der Tatsache, dass bei einem Durchfluss von bestimmter Geschwindigkeit durch eine kritische Verengung hindurch die Strömung von ihrer laminaren Form hinter dieser Enge in eine turbulente, d.h. eine parallele Teilchenbewegung in eine verwirbelte übergeht. Das akustische Resultat dieser Störung des Luftstroms besteht in einer stochastischen (statistischen) Variation des Luftdrucks über die Zeit, einem sog. Rauschen. Derartige Geräuschquellen sind artikulatorisch prinzipiell entlang der gesamten Länge des Ansatzrohres (bei entsprechend aufeinander abstimmbaren Engebildung und Luftstromgeschwindigkeit) von der Glottis bis zu den Mundlippen bildbar. Für die spektrale Charakteristik des so erzeugten Geräusches ist dabei hauptsächlich sowohl der geräuschbildende Engeabschnitt des Ansatzrohres sowie – als Modifikator – der daran anschließende Teil bis hin zu den Mundlippen verantwortlich.

### *Die Filterfunktion des Ansatzrohres*

Den für das Funktionieren der lautsprachlichen Kommunikation wichtigsten Aspekt der Sprachschallproduktion bildet die Modifikation des Rohschalls durch die Filterwirkung des Ansatzrohres.

Akustisch gesehen kann man nach Fant (1960) den Prozess der Sprachschallproduktion durch eine Rohschallquelle und ein Filter modellieren (vgl. Abb. 8 u. 9): Im Zeitbereich lässt sich dieser Prozess als Faltung des glottalen Quellsignals mit der Impulsantwortfunktion des Ansatzrohrhohlraums darstellen (Abb. 8), im Frequenzbereich entspricht dies – wesentlich anschaulicher – der Multiplikation des glottalen Spektrums mit dem des Ansatzrohrfilters (Abb. 9).



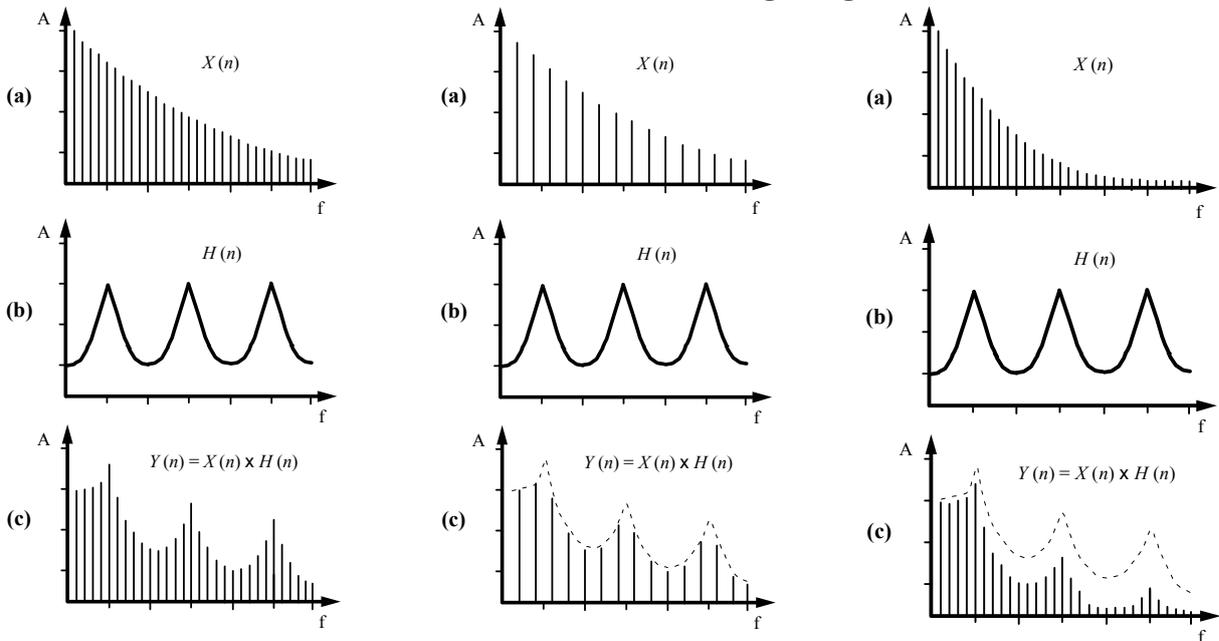
**Abb. 8: Die Quelle-Filter-Theorie: Der Zusammenhang zwischen Quellsignal und Filtercharakteristik im Zeitbereich**

Das Ansatzrohr, das zwischen der Rohschallquelle und den Mundlippen liegt, fungiert in Abhängigkeit von seiner jeweiligen geometrischen Form als variabler Filter bzw. Resonator, d.h., dass von ihm bestimmte Frequenzen – seine Eigen- bzw. Resonanzfrequenzen – verstärkt durchgelassen, andere Frequenzen des Eingangssignals jedoch gedämpft werden (vgl. die ‚Gipfel‘ bzw. ‚Täler‘ in Abb. 9 (b)). Die Resonanzfrequenzen sind dadurch gekennzeichnet, dass sie im Ansatzrohr stehende Wellen bilden. Die Lage der Resonanzfrequenzen bestimmt sich dabei allein durch den Gesamtverlauf der Querschnittsfläche des Rohres entlang seiner Mittelachse. Dieser Zusammenhang ist durch die Webstersche Hornleichung – eine partielle Differentialgleichung zweiten Grades – beschreibbar. Es soll hier auf die komplexe Mathematik dieses Zusammenhangs nicht näher eingegangen werden. Wir wollen vielmehr versuchen, diese Filterwirkung des Ansatzrohres am Reflexionsverhalten im neutralen Rohr zu veranschaulichen.

### Das Modell des neutralen Rohres

Für unsere Ausgangsüberlegung stellen wir uns das Ansatzrohr als gerades kreiszylindrisches Rohr von 17 cm Länge und einem gegenüber seiner Länge relativ geringen Durchmesser (von maximal 4 bis 5 cm) vor. Dieses – entlang seines Gesamtverlaufs  $z$  eine gleichbleibende Querschnittsfläche – aufweisende (eben neutrale) – Rohr besitze schallharte Wände (d.h. an denen totale Reflexion auftritt) und sei am einen – der Position der Glottis entsprechenden – Ende ebenso schallhart verschlossen, am anderen Ende – dem der Lippen – hingegen offen.

Tritt in diesem System ein Impuls auf – und breitet sich diese akustische ‚Störung‘ kugelförmig aus –, so werden die Luftmoleküle an den Wänden und am verschlossenen Ende des Rohres total reflektiert, während ihnen am offenen Rohrende kein Reflexionswiderstand entgegengesetzt wird. Diese Tatsachen stellen die Randbedingungen für die sich als **Resonanzen** in diesem System ausbildenden **stehenden Wellen** dar: Am verschlossenen Rohrende ist wegen der totalen Reflexion der Schalldruckwechsel maximal, am offenen Ende hingegen herrscht immer der normale Umgebungsluftdruck (Punkte mit der letztgenannten Eigenschaft nennt man auch ‚Knoten‘ einer stehenden Welle). Auf Grund dieser Randbedingungen bilden sich im neutralen Rohr stehende Wellen nach dem in Abb. 10 gezeigten Schema aus:



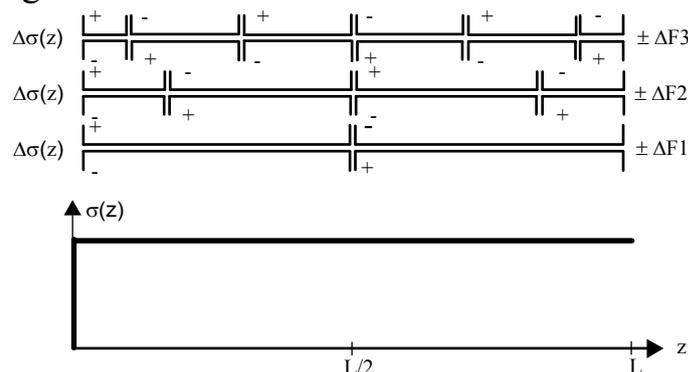
**Abb. 9: Der Zusammenhang zwischen Quallsignal und Filtercharakteristik im Frequenzbereich: (a) Linienspektrum des Rohrschallsignals, (b) Spektrum der Impulsantwort (Filterfunktion), (c) Linienspektrum des Ausgangssignals. Mitte: Quallsignal (a) von doppelter Grundfrequenz gegenüber links (mit unter (c) punktiert eingezeichneter spektraler Hüllkurve); rechts: stärker gedämpftes Quallsignal (mit unter (c) punktiert eingezeichneter spektraler Hüllkurve des ursprünglichen Beispiels links)**

Die frequenzmäßig erste Welle, die diesen Randbedingungen genügt, ist ein Cosinus mit einer Wellenlänge von viermal der Rohrlänge  $L$ , die zweite Resonanzfrequenz passt bezüglich ihrer Wellenlänge zu Dreiviertel (mit einem weiteren Knoten)



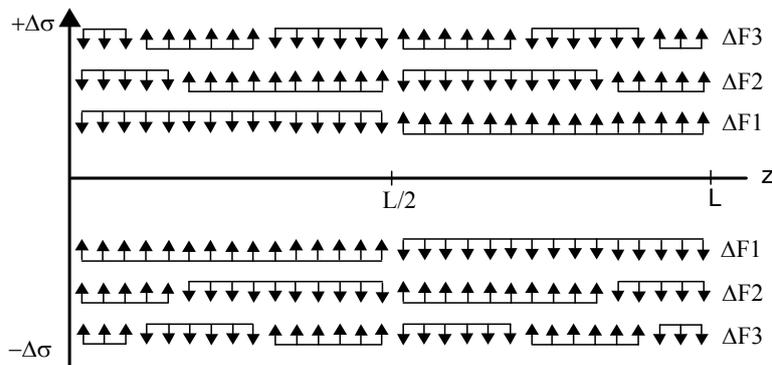
Beim natürlichen menschlichen Ansatzrohr haben wir natürlich nicht einen verlustfrei reflektierenden Körper vor uns, und auch das hintere (glottale) Ende ist nicht zu jedem Zeitpunkt verschlossen. Die Auswirkung hiervon ist, dass wir es im Fall des natürlichen Vokalschalls nicht mit einzelnen Resonanzfrequenzen sondern mit Frequenzbändern, die verstärkte Energie aufweisen – den sog. **Formanten** – zu tun haben. Diese Formanten werden im Sonagramm als Balken stärkeren Schwärzungsgrades sichtbar. Neben ihrer **Frequenzlage** sind die natürlichen Vokalformanten zusätzlich durch ihre **Bandbreite** gekennzeichnet. (Letztere ist definiert als der Frequenzbereich [in Hz], den der Formant bei -3dB unter seiner Spitzenamplitude in der Breite einnimmt.)

Die Resonanzen des nicht querschnittsflächengleichen Rohres lassen sich nicht in gleicher Weise wie beim neutralen Rohr veranschaulichen. Der Einfluss lokaler Querschnittsveränderungen auf die Frequenzlage der einzelnen Resonanz ist abhängig von der Energieverteilung der entsprechenden stehenden Wellen im neutralen Rohr (dem Produkt aus deren Schalldruck- und Schallschnelleverteilung) entlang der Rohrachse. Generell zeigt sich ein Bild, wie es Tillmann (1980) am Modell der Formantverschieber (vgl. Abb. 11) darstellt: Eine Verengung der vorderen, lippenseitigen Hälfte des Ansatzrohres hat dieselbe Wirkung wie eine entsprechende Erweiterung der hinteren, glottisseitigen Hälfte auf die Lage der ersten Resonanz, nämlich eine Frequenzabsenkung.



**Abb. 11: Die Formantverschieber nach Tillmann (1980): Die Unterteilung des Ansatzrohres in seiner Länge bezüglich der frequenzerhöhenden/-erniedrigenden Wirkung ( $\Delta F_N$ ; durch Vorzeichen gekennzeichnet) von lokalen Verengungen/Erweiterungen ( $\Delta\sigma$ ; durch unterschiedlich lange, nach oben bzw. unten offene Haken schematisiert) für die einzelnen Resonanzfrequenzen (FN)**

Bezüglich der höheren Resonanzen – entsprechend der Energieverteilung der zugehörigen stehenden Wellen des neutralen Rohres – sind die relevanten Bereiche der Ansatzrohrachse dementsprechend anders unterteilt: für die zweite Resonanz zeigt sich ein gleicher – frequenzvermindernder – Einfluss einer Verengung im vorderen Sechstel wie einer Erweiterung im hintersten Sechstel des Ansatzrohres; ebenso ist der Einfluss einer Verengung des Drittels vor der Ansatzrohrmitte in seiner frequenzerhöhenden Wirkung gleich der Erweiterung des rückwärts daran anschließenden Drittels. In der Abb. 12 (nach Tillmann 1980) sind die nach dem kleinsten gemeinsamen Vielfachen der Ansatzrohrunterteilung (30) möglichen Formantverschiebungen für die drei ersten Resonanzfrequenzen dargestellt.

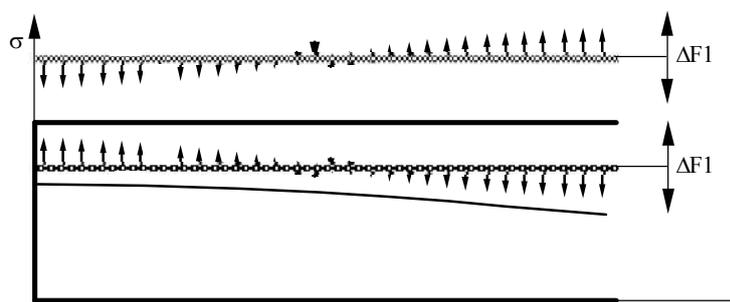


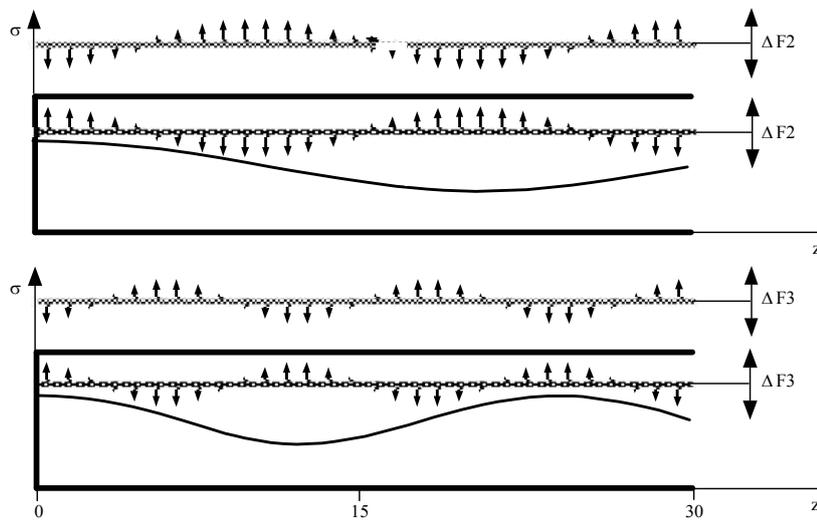
**Abb. 12: Das nach dem kleinsten gemeinsamen Vielfachen der Formantverschieberlänge unterteilte Ansatzrohr: Formantfrequenzveränderung (Pfeile) auf Grund lokaler Ansatzrohrerweiterung (obere Hälfte) bzw. Ansatzrohrverengung (untere Hälfte) (nach Tillmann 1980)**

Hier soll hervorgehoben sein, dass sich das akustische Vokalsystem durch dieses Modell als Abbildung der geometrischen Veränderungen gegenüber dem neutralen Rohr in die Verschiebung der Resonanzen aus der Neutrallage verstehen lässt.

Am Schema der Formantverschieber lässt sich eine weitere akustisch-phonetische Tatsache klarmachen. Auf Grund des bezüglich des Vorzeichens entlang der Rohrachse ständig wechselnden Einflusses geometrischer Veränderungen auf die Resonanzfrequenzen ergibt sich, dass die Abbildung der artikulatorischen Geometrie in die spektralen Eigenschaften des resultierenden Sprachschalls zwar eindeutig, nicht aber ein-eindeutig ist: Ein Schall gegebener spektraler Komposition kann durchaus durch sehr unterschiedliche geometrische Ansatzrohrkonfigurationen hervorgebracht werden. Abb. 12 mag auch verdeutlichen, wieso im allgemeinen die Betrachtung der ersten drei Formanten für die phonetische Analyse ausreichend ist: Die willkürliche Beeinflussung der Lage des dritten Formanten ist schon an so eng umschriebene lokale Veränderungen des Ansatzrohres gebunden, dass diese lediglich bei den feinmotorisch komplexen konsonantischen Artikulationen vorkommen (vgl. u.). Globalere Zungenlageveränderungen heben sich in ihrer Wirkung auf die Lage der höheren Resonanzen wegen ihres wechselnden Einflusses gegenseitig auf: Eine gezielte, willkürliche Beeinflussung ist ab dem vierten Formanten (bei dem die gleiche geometrische Änderung alle 2,3 cm entlang des Ansatzrohres jeweils den gegenteiligen Effekt auslösen würde) somit nicht möglich. Ab dem vierten Formanten ist deren Frequenz eher abhängig von der je individuellen Ansatzrohrgeometrie, weniger der vorliegenden Artikulation.

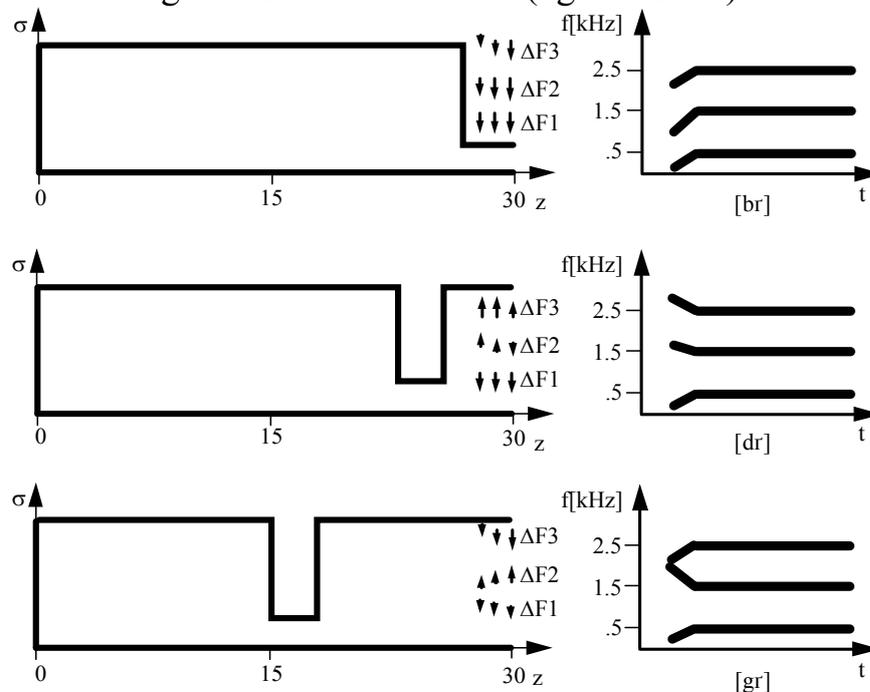
Wie bereits ausgeführt, steht der Einfluss einer lokalen Querschnittsveränderung in Beziehung zu der sich wiederum sinoidal entlang der Rohrachse ändernden Energieverteilung der stehenden Welle im neutralen Rohr. In Abb. 13 ist so (in gleicher Unterteilung wie in Abb. 12) der Frequenzverschiebungseffekt lokaler Änderungen auch **quantitativ** anhand der dargestellten Pfeillänge gekennzeichnet.



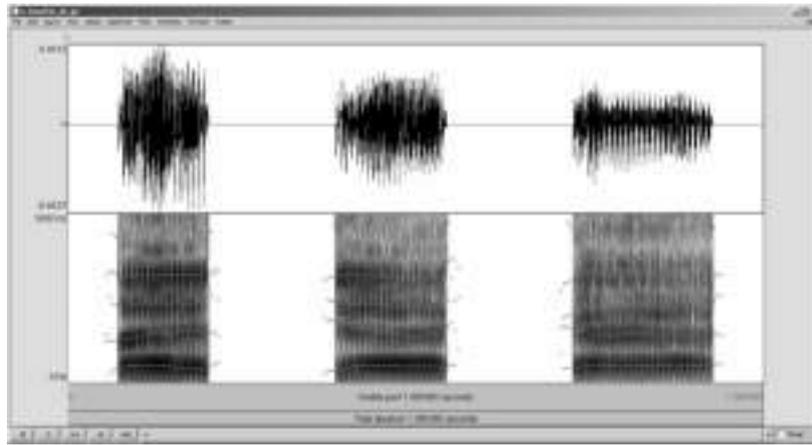


**Abb. 13: Die abgestufte Wirkung lokaler Querschnittsflächenänderung gegenüber dem neutralen Rohr (schwarz, mit den eingezeichneten stehenden Wellen): bei Vergrößerung (nach oben, hellgrau); Verkleinerung (nach unten, dunkelgrau) als Frequenzverschiebung (Pfeile) gegenüber dem Neutralvokal [e]**

Die Abb. 14 schließlich zeigt das sich hieraus ergebende Formantbewegungsmuster (Transition), das sich ergibt, wenn wir von einer labialen, alveolaren bzw. velaren Ansatzrohrverengung ausgehend wieder in eine neutrale Ansatzrohrstellung zurückkehren. Die in der rechten Hälfte der Abbildung dargestellten resultierenden Formanttransitionen entsprechen nun tatsächlich denen, wie wir sie bei natürlich-sprachlichen Silbenäußerungen beobachten können (vgl. Abb. 15).

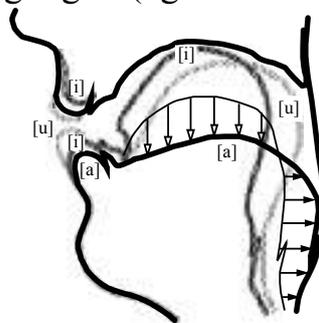


**Abb. 14: Durch den Übergang von der links mit ihrer Formantverschiebungswirkung (Pfeile) dargestellten Ansatzrohrkonfiguration zur Neutralposition bewirkten Formanttransitionen (rechte Seite)**

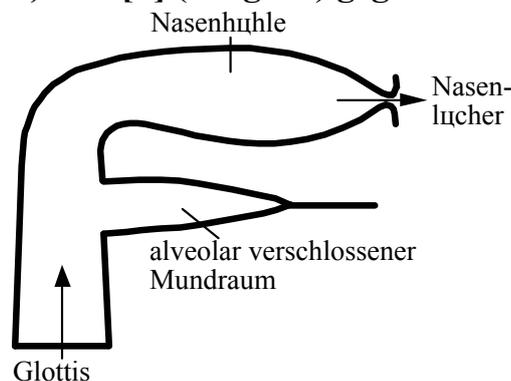


**Abb. 15: Sonagramme der Silben [bə], [də], [gə]**

Die Formantlage der Eckvokale [a], [i] und [u] – im Kontrast zum Neutralvokal [ɤ] – wollen wir uns abschließend anhand der Abb. 16 klarmachen: Beim [a] haben wir es wie durch die Pfeile in der Abbildung gekennzeichnet mit einer Erweiterung der vorderen Hälfte des Ansatzrohres und einer Verengung der hinteren Hälfte des Ansatzrohres zu tun. Dies führt beides (vgl. Abb. 11) zu einer Anhebung des ersten Formanten, während sich die Wirkungen bezüglich der Lage des zweiten Formanten faktisch aufheben. Beim [i] und [u] sind die Verhältnisse der Flächenveränderung der beiden für den ersten Formanten maßgeblichen Einzelhälften gerade umgekehrt, d.h. der ersten Formant wird gesenkt. Die lokale Ansatzrohrverengung liegt jedoch so, dass sie beim [i] (hauptsächlich in den hinteren beiden Drittel der vorderen Ansatzrohrhälfte) zusammen mit der Erweiterung der hinteren Ansatzrohrhälfte eine Erhöhung des zweiten Formanten bewirkt, während sie beim [u] (hinten in den beiden vorderen Dritteln) eine absenkende Wirkung ergibt (vgl. a. Abb. 18).



**Abb. 16: Schematische Sagittaldarstellung der Artikulation der Eckvokale [a] (schwarz), [i] (dunkelgrau) und [u] (hellgrau) gegenüber der Neutrallage (dünn)**



**Abb. 17: Schema der Ansatzrohrkonfiguration für die Produktion eines alveolaren Nasals**

Wird das Velum gesenkt und somit der Nasenraum zugeschaltet, so haben wir es mit einem vom Rachenraum an nach vorne zweigeteilten Ansatzrohr zu tun. Bei einem Röhrensystem dieser Art kommt es zu Interferenzen zwischen den Resonanzen der Teilsysteme, wie sie anhand der obigen Abb. 17 zur Nasalproduktion verdeutlicht sind: Als Hauptansatzrohr fungiert hier der Rachen- und Nasaltrakt, der eine erste Resonanz bei ca. 250 Hz aufweist und die höheren Frequenzen stark dämpft. Hinzu kommen die Resonanzeigenschaften des Mundraumes als geschlossenem Seitenrohr, die in einer solchen Konstellation als Antiresonanzen wirksam werden und im Spektrum als Intensitätseinbrüche auszumachen sind. Die frequenzmäßige Lage dieser Antiresonanzen ist wiederum von der Länge und der geometrischen Gestalt dieses Rohres abhängig. Wenn wir von einem neutralen Rohr einer Länge von 7 cm<sup>18</sup> ausgehen, so entspräche die erste dieser Frequenzen etwa 1200 Hz. Messungen an natürlich gesprochenem Material ergaben Antiresonanzen bei [m] zwischen 750 und 1250 Hz, bei [n] zwischen 1450 und 2200 Hz und bei [ŋ] über 3000 Hz.

#### 4. Praktische Lautsegmentation mit Hilfe des Sonagramms

Im Folgenden sollen in aller Kürze die wesentlichen sonographischen Muster der unterschiedlichen Sprachlautkategorien dargestellt und zudem die in der phonetischen Forschung gängigen Kriterien der Segmentation anhand dieser Muster vorgestellt werden.

Wie in den vorangegangenen Abschnitten verdeutlicht, ergeben sich die sonographischen Muster aus den Spektraleigenschaften des Quellsignals und denen des Ansatzrohrfilters. Bei den stimmhaften Sonoranten werden wir es also mit aus hervorgehobenen Harmonischen des Rohschalls geformten Formantstrukturen zu tun haben, bei stimmlosen Frikativen mit kontinuierlichen Spektren in Abhängigkeit von der ab der geräuschverursachenden Verengung bis zu den Lippen verbleibenden Länge des Ansatzrohres. Da bei der natürlichen Artikulation sich die "Sprechwerkzeuge" ja kontinuierlich bewegen und sich somit auch die akustischen Eigenschaften des Ansatzrohres kontinuierlich verändern, sind Einzellaute physikalisch nicht klar voneinander abgegrenzt, sondern gehen ineinander über, so dass wir uns bei der Segmentation auf operationalisierbare Kriterien beziehen müssen. Wie wir im Folgenden sehen werden, bietet uns das Sonagramm hierfür die wesentlichen Hilfestellungen.

#### *Vokale*

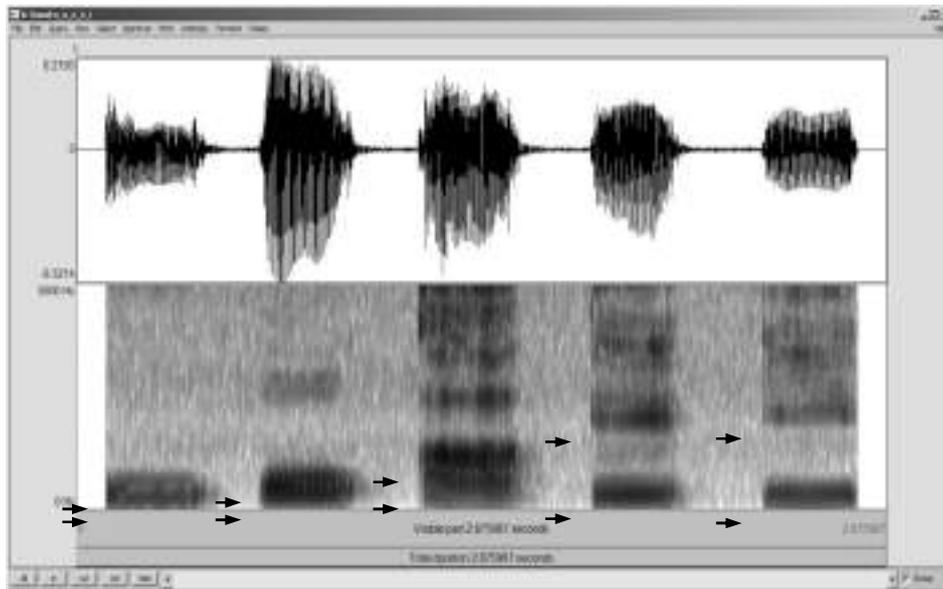
Wie bereits im letzten Abschnitt genauer ausgeführt, sind die einzelnen Vokale durch die Frequenzlage ihrer Formanten voneinander unterschieden, wobei die wesentliche Rolle den beiden untersten Formanten zukommt (vgl. Abb. 18 u. 19). In erster Näherung bildet die erste Formantfrequenz den Öffnungsgrad des Vokals ab und die Frequenz des zweiten Formanten die horizontale Zungenlage (vorne = hoch).

In Abb. 18 ist dasselbe Signal unter (a) als das übliche Breitband-sonagramm, bei dem die Formanten (F1 und F2 durch Pfeile markiert) erkennbar werden (PRAAT *Spectrogram settings: Window length 0.005*), dargestellt, unter (b) als Schmalbandsonagramm mit visueller Auflösung der einzelnen Harmonischen (*Window length 0.03*)

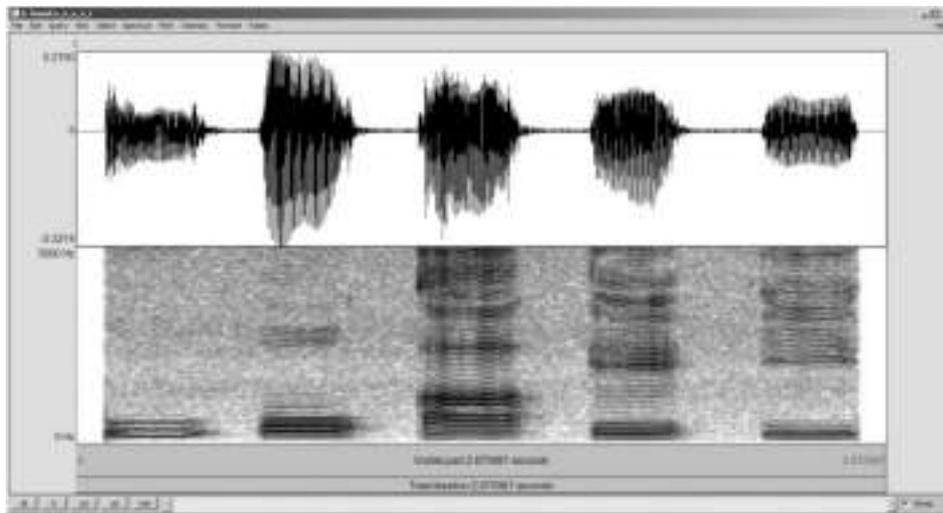
---

<sup>18</sup>Für den oralen Teil des Ansatzrohres.

(a)

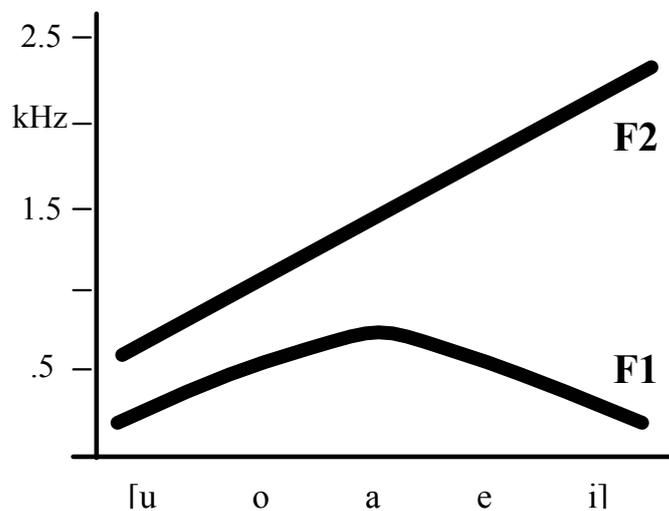


(b)



**Abb. 18: PRAAT-Darstellung der Vokale [u], [o], [a], [e] und [i]:  
(a) Breitband-, (b) Schmalbandsonagramm**

Abb. 19 fasst dies nochmals als Merkschema zusammen.

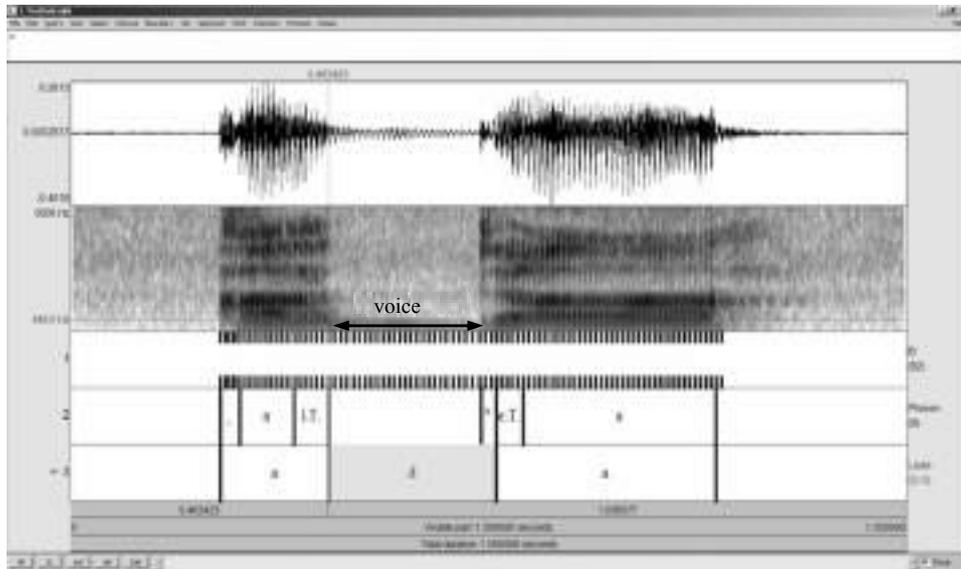


**Abb. 19: Merkschema für Vokalformantfrequenzen.**

### *Konsonanten: Plosive*

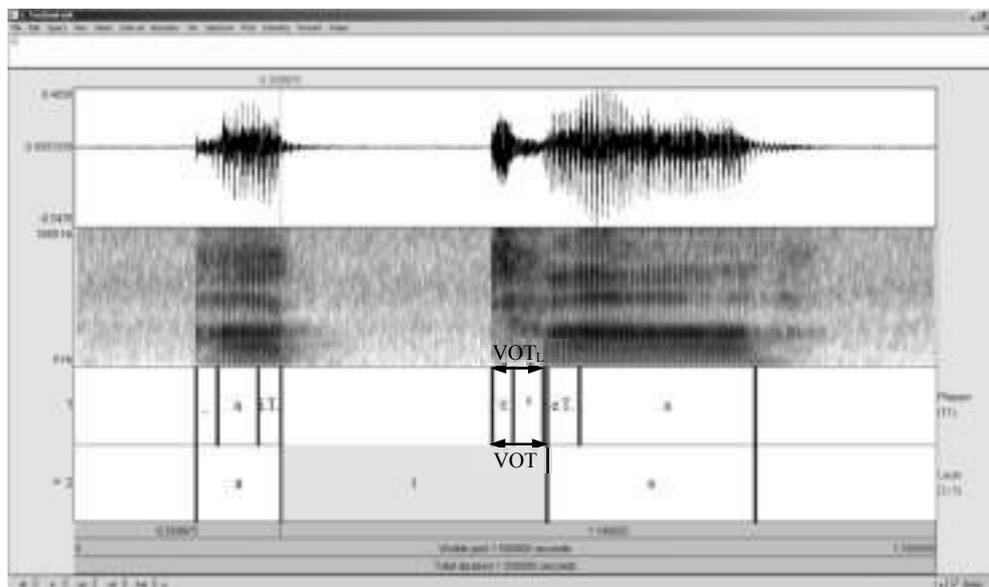
Am Beispiel der Verschlusslaute seien nun zuerst die bisher nicht besprochenen im Breitbandsonogramm erkennbaren Merkmale und die artikulatorisch-akustischen Phasen der Konsonanten erläutert.

Abb. 20 zeigt eine PRAAT-Annotation der Äußerung [ʔa'da:] bezüglich der Startzeitpunkte der einzelnen Glottisimpulse (1) sowie bezüglich der Abgrenzung einzelner Phasen (2) und von Einzellaute (3).



**Abb. 20: PRAAT-Annotation der Äußerung [ʔa'da:]**

Abb. 21 zeigt das Parallelbild für die Äußerung [ʔa'tʰa:] mit der Segmentation von einzelnen Phasen (1) und Einzellaute (2)

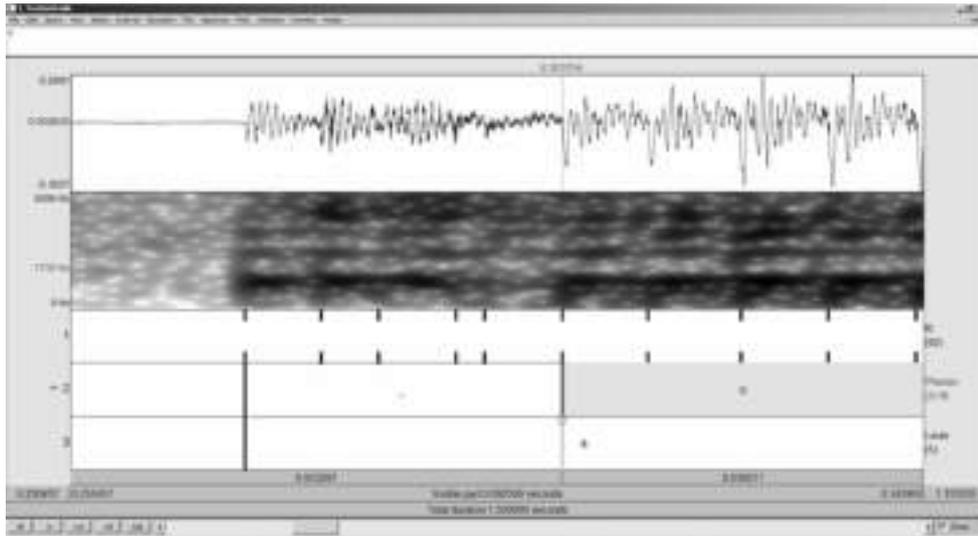


**Abb. 20: PRAAT-Annotation der Äußerung [ʔa'tʰa:]**

Die Segmentationsmarkierungen in den obigen Abbildungen 20 und 21 sind in den periodischen Signalabschnitten immer am Zeitpunkt des Nulldurchgangs (hier negativ) vor der Führungsamplitude der Folgeperiode orientiert. Dies ist die

gängige Praxis bei der phonetischen Lautsegmentation, woraus sich auch eine generelle Fehlertoleranz von  $\pm$  einer Stimmtoneperiode (ca.  $\pm$  10 ms bei Männerstimme) ergibt.

Die erste Segmentationsebene der Abb. 20, die in Abb. 22 nochmals vergrößert dargestellt ist, zeigt so die Aufeinanderfolge einzelner Stimmtoneperioden, die bezüglich des Signaleinsatzes zuerst unregelmäßig erscheint, um ab der Cursor-Position in Abb. 22 dann regelmäßig zu werden. Die erste – mit [ˌ] annotierte – Phase entspricht dabei dem in enger Transkription mit [ʔ] wiedergegebenen und hier durch Knarrstimme realisierten phonetischen Segment, das in der Lautsegmentation hier dem ersten [a] zugerechnet wird.



**Abb. 22: Vergrößerter Anfangsausschnitt von Abb. 20**

Die Markierungen der Ebene 1 weichen in Abb. 20 auch am Signalende von den Segmentationen der Phasen und Lauten ab: Während im Oszillogramm noch (die unter (1) segmentierte) Periodizität sichtbar ist, sind im Sonagramm ab der auf Ebene 2 und 3 gesetzten Endgrenzen weder F2 noch höhere Formanten erkennbar, was für die phonetische Lautsegmentation als Hinweis für das Vokalende gilt.

In Abb. 20 und 21 sind zudem die weiteren durch die Konsonantenartikulation bedingten akustischen Phasen markiert. Hierbei handelt es sich – in zeitlicher Abfolge – um (a) die durch die in ihrer Richtung artikulationsstellenspezifischen Formantbewegungen (vgl. Abb. 14 u. 15) gekennzeichnete “implosive” Transition (in den Abbildungen als “i.T.” abgekürzt), die durch die artikulatorische Enge-/Verschlussbildung verursacht wird, (b) die – in Abb. 20 und 21 nicht annotierte – Konstriktions- (hier Verschluss-) Phase, (c) die ggf. weiter unterteilbare geräuschhafte Verschlusslösungsphase und (d) die wiederum artikulationsstellenspezifische “explosive” Transition (als “e.T.” abgekürzt), die durch die Öffnungsbewegung zum Folgevokal verursacht ist.

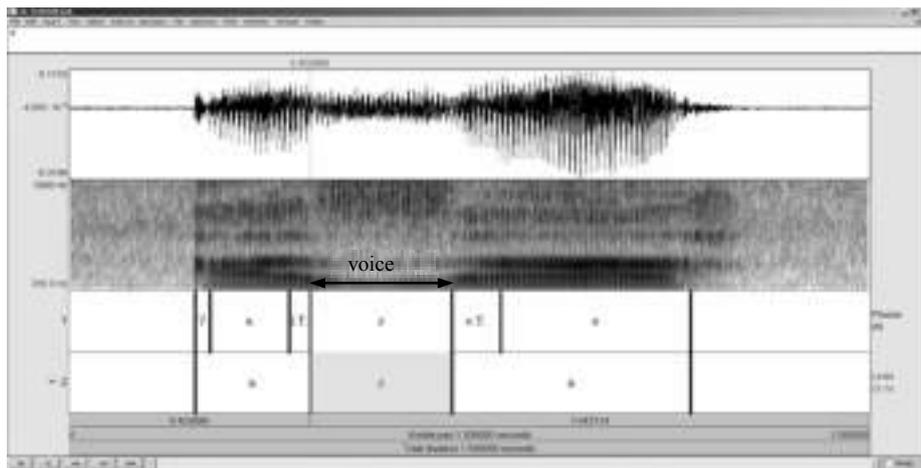
Bei der Lautabgrenzung werden die Plosiv-Transitionen bis zum Verschwinden bzw. ab dem Einsetzen von F2 und höheren Formanten zu den Umgebungsvokalen gezählt, während als Plosivdauer die Verschluss- plus die geräuschhafte Verschlusslösungsphase gilt.

Die Verschlussphase des stimmhaften Plosivs [d] in Abb. 20 ist durch eine anhaltende Periodizität gekennzeichnet, die im Sonagramm an der sog. “voice bar” erkennbar ist. Bei [t] hingegen (Abb. 21) zeigt sich hier eine echte Signalpause.

Die Verschlusslösungsphase bei [d] (Abb. 20) zeigt sich in einer kurzzeitigen Geräuschüberlagerung, während sie bei [t] zwei spektral unterschiedliche Geräuschabschnitte umfasst: das Verschlusslösungsgeräusch (“burst” – an der Artikulationsstelle gebildet; als [t] annotiert) und die darauf folgende (glottale) Behauchung oder Aspiration [h], die zusammen die sog. “voice onset time” (nach Lisker/ Abramson bis zur ersten Periode ( $VOT_{LA}$ ), nach Klatt bis zum F2-Einsatz ( $VOT_K$ )) bilden.

#### *Frikative*

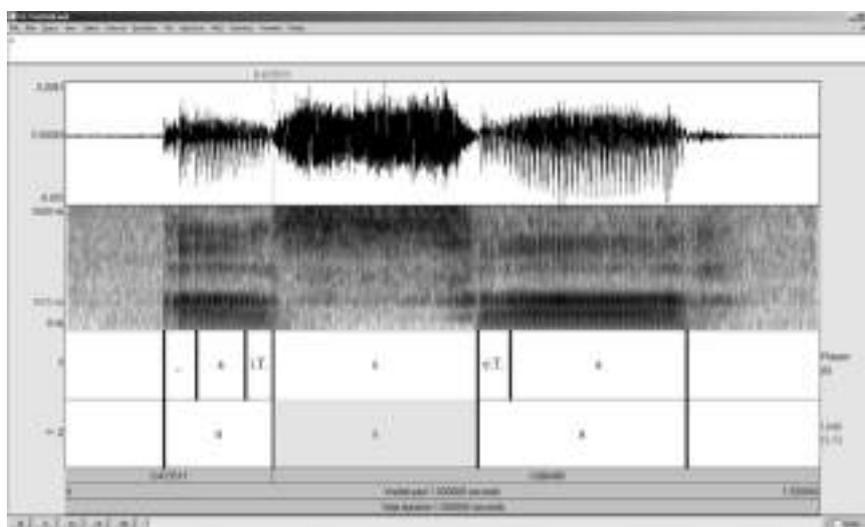
Abb. 23 und 24 geben parallel zu oben die entsprechenden Segmentationen für die artikulationsstellengleichen Frikative anhand der Äußerungen [ʔa'za:] und [ʔasa:] wieder.



**Abb. 23: PRAAT-Annotation der Äußerung [ʔa'za:]**

Bei den Frikativen zeigen sich auch im Zeitverhalten ähnlich gestaltete artikulationsbedingte Phasen wie bei den Plosiven, wobei während der Konstriktionsphase das frikativische Geräuschsignal auftritt. Auch während dieser Phase sind zusätzliche spektrale Veränderungen (Transitionen) sichtbar, die sich aber schwer segmentieren lassen.

Auch bei den Frikativen werden die Transitionen üblicherweise der Dauer der flankierenden Vokale zugeschlagen. [Wichtiger praktischer Hinweis: Hört man nur das vokalische Segment inklusive Transitionen ab, so ergibt sich aufgrund deren artikulationsstellenspezifischer Form als Hörartefakt eine Silbe mit homorganem Plosiv.]



**Abb. 24: PRAAT-Annotation der Äußerung [ʔa'sa:]**

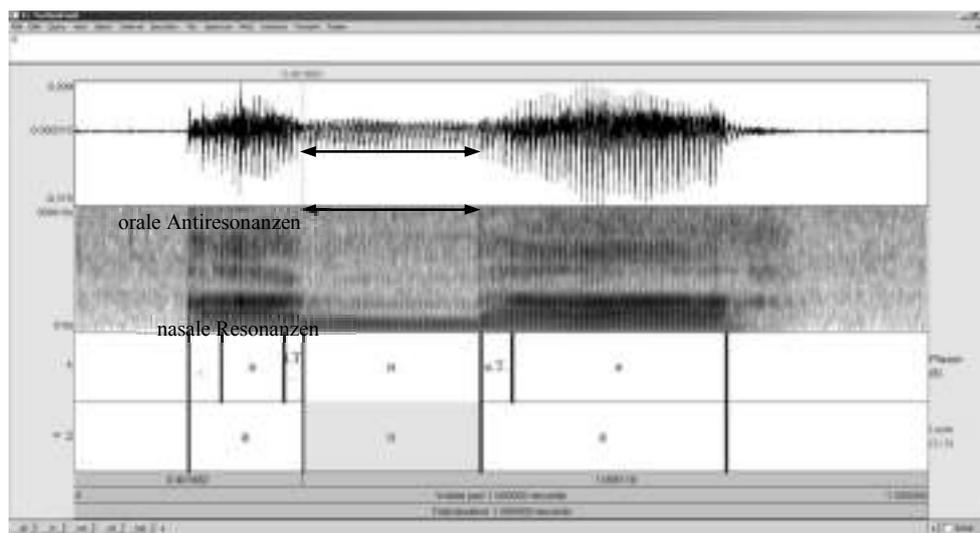
Das akustische Hauptunterscheidungsmerkmal bei den Frikativen ist die spektrale Lage und die Ausprägung des Geräusches während der Konstriktionsphase. Die spektrale Lage ist hierbei vor allem abhängig von der Länge des Ansatzrohres zwischen der geräuschverursachenden Enge bis zu den Mundlippen, die sog. “artikulatorische Tiefe” (vgl. Abb. 25). Intensitätsmäßig besonders ausgeprägt sind die sog. Sibilanten (alveolare und postalveolare Frikative), bei denen eine doppelte Verwirbelung des Luftstroms stattfindet: einmal primär an der Engstelle und danach zusätzlich an den Schneidezähnen.



**Abb. 25: Die spektrale Ausprägung der “artikulatorischen Tiefe” am Beispiel der deutschen Wörter *Asse, Asche, Ache***

#### *Nasale*

Abb. 26 zeigt die Segmentation der Äußerung [ʔa'na:] mit der aufgrund der Velumsenkung durch nasale Resonanzen und orale Antiresonanzen (und hierdurch teils verkürzte Transitionen) gekennzeichneten Konstriktionsphase.

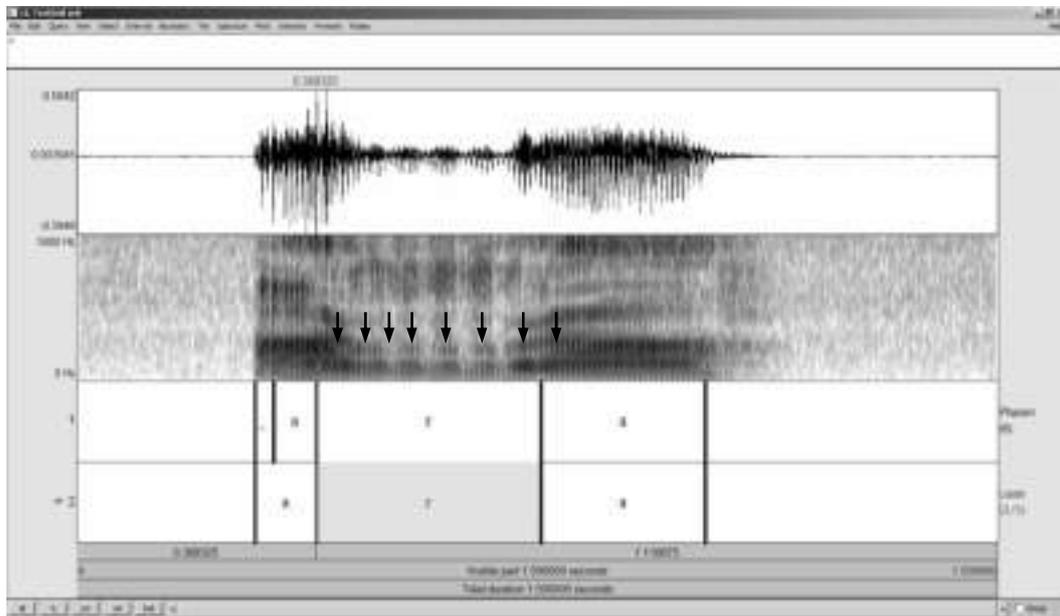


**Abb. 26: PRAAT-Annotation der Äußerung [ʔa'na:]**

#### *Gerollte Laute*

Abb. 27 zeigt die Segmentation der Äußerung [ʔa'ra:] mit stark gerolltem Zungenspitzen-*r*, die während der konsonantischen Konstriktionsphase im Sonagramm die

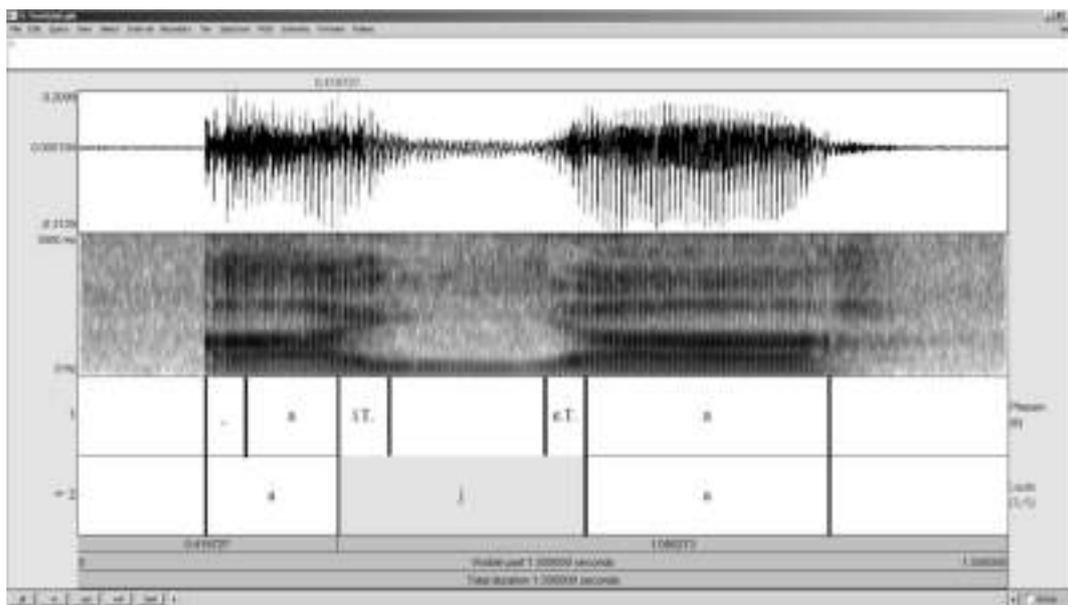
mit Pfeilen markierten wiederholten kurzzeitigen Verschlüsse aufweist. Die Transitionsphasen sind ebenso durch diese Verschlüsse gekennzeichnet und zählen somit auch zum konsonantischen Segment.



**Abb. 27: PRAAT-Annotation der Äußerung [ʔa'ra:]**

*Gleitlaute*

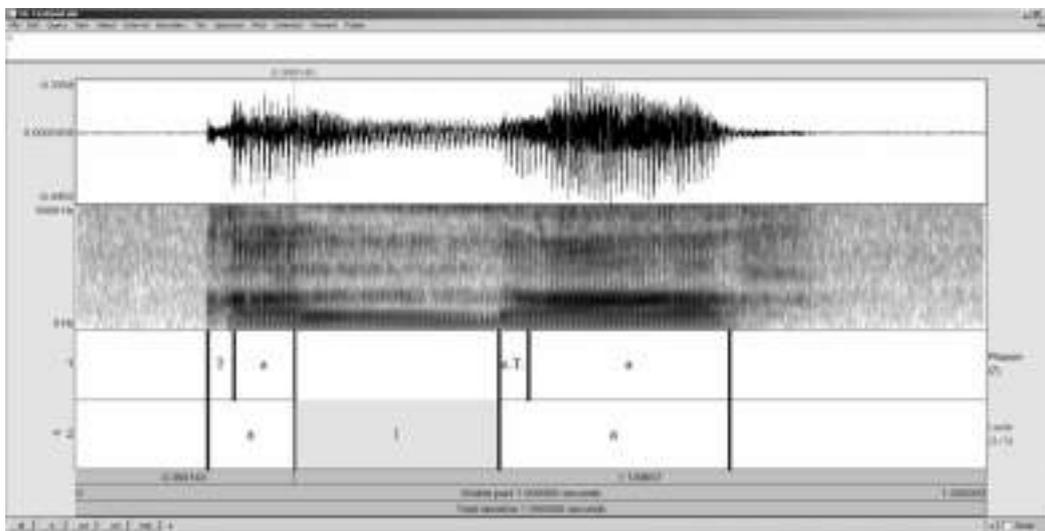
Abb. 28 zeigt die Segmentation der Äußerung [ʔa'ja:] die im Sonogramm durch in der Richtung den bisherigen homorganen Konsonanten entsprechenden doch wesentlich länger andauernden Transitionsphasen gekennzeichnet ist. Diese werden hier größten Teils der konsonantischen Dauer zugeschlagen und sind dabei nur schwer, nicht rein visuell und nur unter periodenweiser auditiver Kontrolle segmentierbar.



**Abb. 28: PRAAT-Annotation der Äußerung [ʔa'ja:]**

### Laterale

Was zuletzt zu den Segmentationschwierigkeiten bei Gleitlauten ausgeführt wurde, gilt in entsprechender Form auch für Laterale wie sie in Abb. 29 anhand der Äußerung [ʔa'la:] dargestellt sind.



**Abb. 29: PRAAT-Annotation der Äußerung [ʔa'la:]**

### 5. Epilog: Segmentation fließend gesprochenen Sprache

Für die Segmentation fließend gesprochenen Sprache bietet sich wegen der hierbei häufig auftretenden Reduktionserscheinungen eine Mehrebenen-Segmentation/Annotation an. So sollten einerseits – soweit möglich – alle Einzelwörter abgegrenzt und bezüglich ihrer kanonischen Aussprache annotiert, deren realisierte Aussprache zusätzlich phonetisch eng transkribiert und – je nach Aufgabenstellung – lautbezogen segmentiert werden.

Die folgende Abbildung soll dies – ohne weitere Kommentierung – an einem Ausschnitt der Äsopfabel von “Nordwind und Sonne” in ukrainischer Sprache (vgl. Pompino-Marschall/ Steriopolo/ Žygis, im Druck) exemplarisch verdeutlichen.

In Abb. 30 ist das Ende der Fabel dargestellt, das in breiter phonologischer Transkription folgendermaßen lautet:

[...]  $\text{ʃ}^{\text{h}}\text{ʃ}^{\text{h}} \text{'s}^{\text{h}}\text{ɔ}^{\text{h}}\text{n}^{\text{h}}\text{t}^{\text{h}}\text{s}^{\text{h}} \text{s} \text{'p}^{\text{h}}\text{ɔ}^{\text{h}}\text{m}^{\text{h}}\text{ɪ}^{\text{h}}\text{z} \text{n}^{\text{h}}\text{ɪ}^{\text{h}} \text{'d}^{\text{h}}\text{v}^{\text{h}}\text{ɔ}^{\text{h}} \mid \text{b}^{\text{h}}\text{u}'\text{l}^{\text{h}}\text{ɔ} \text{s}^{\text{h}}\text{l}^{\text{h}}\text{'n}^{\text{h}}\text{i}^{\text{h}}\text{ʃ}^{\text{h}}\text{ɪ}^{\text{h}}\text{m}.$

Dem steht die Realisation als

[...]  $\text{ʃ}^{\text{h}} \text{'s}^{\text{h}}\text{ɔ}^{\text{h}}\text{n}^{\text{h}}\text{t}^{\text{h}}\text{s}^{\text{h}} \text{'s}^{\text{h}}\text{p}^{\text{h}}\text{ɔ}^{\text{h}}\text{m}^{\text{h}}\text{ɪ}^{\text{h}}\text{z} \text{n}^{\text{h}}\text{ɪ}^{\text{h}} \text{'d}^{\text{h}}\text{v}^{\text{h}}\text{ɔ}^{\text{h}} \mid \text{b}^{\text{h}}\text{u}'\text{t}^{\text{h}} \text{s}^{\text{h}}\text{l}^{\text{h}}\text{'n}^{\text{h}}\text{i}^{\text{h}}\text{ʃ}^{\text{h}}\text{ɪ}^{\text{h}}\text{m}$   
gegenüber.

Abschließend sei hierfür noch die orthographische Form und die Transliteration der Textstelle gegeben.

[...] *що Сонце з поміж них двох було сильнішим.*

[...] *ščo Sonce z pomiž nyx dvox bulo syl'nišym.*

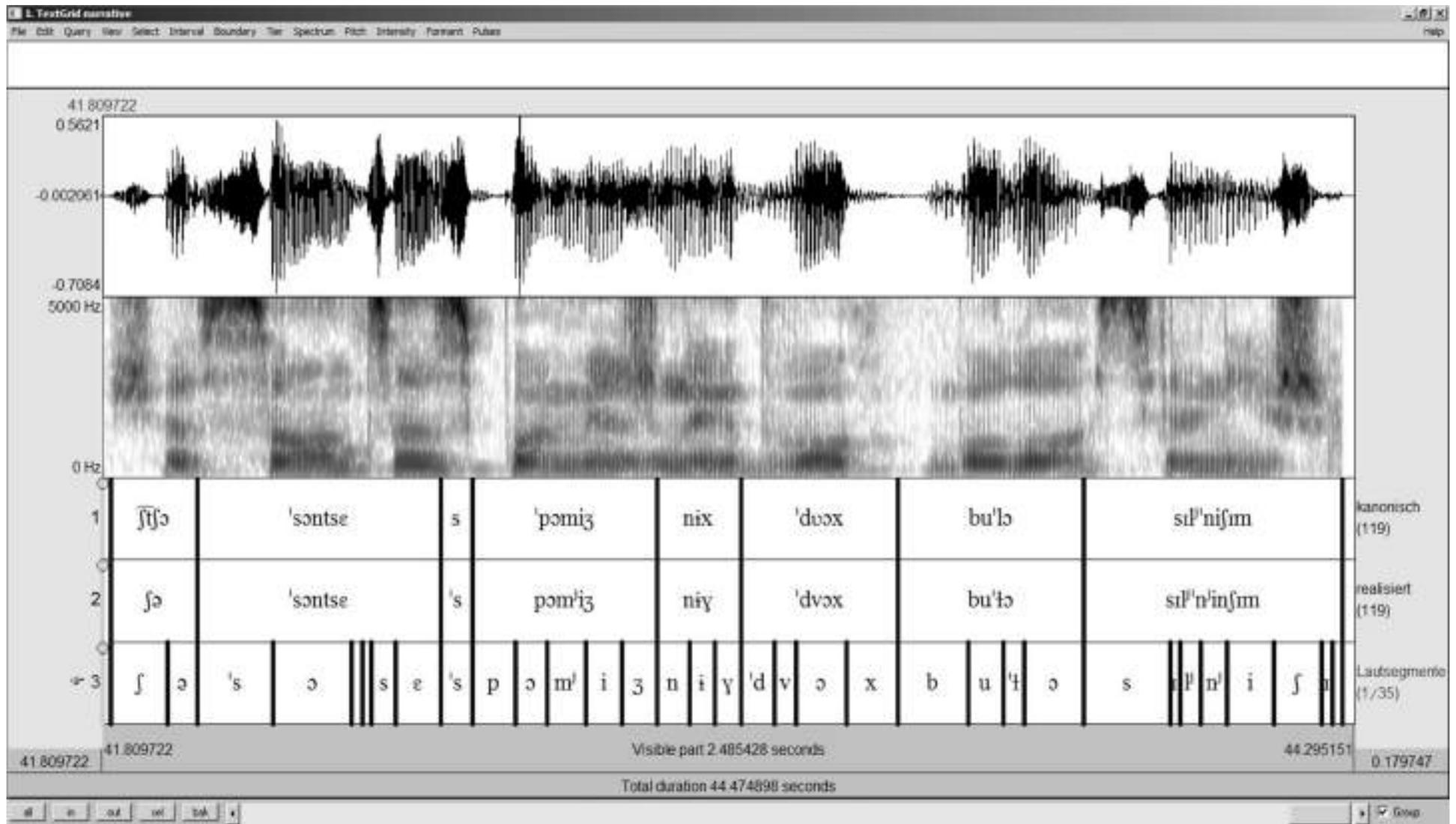


Abb. 30: PRAAT-Annotation der Realisierung des Endes der Fabel von “Nordwind und Sommer” in ukrainischer Sprache; kanonische und realisierte Wortsegmentation und -Annotation sowie Einzellautessegmentation

## LITERATUR

1. Benguerel A.-P. Phonetics, descriptive acoustic. In: Asher, R.E. & Simpson, J.M.Y. (eds.) // The Encyclopedia of Language and Linguistics. – Oxford, 1993. – PP. 3070–3082.
2. Beranek L. Acoustic Measurements. – New York, 194.
3. Boersma P., Weenink D. Praat: doing phonetics by computer [Computerprogramm]. Version 6.0.19 / Boersma Paul, Weenink, David. – 2016 / unter <http://www.praat.org/> [letzter Zugriff am 25.07.2016].
4. Chiba T., Kajiyama M. The Vowel, its Nature and Structure. – Tokyo, 1941.
5. Fant C. G. M. Acoustic Theory of Speech Production.- den Haag, 1960.
6. Fant G. Sound spectrography // Proceedings of the Fourth International Congress of Phonetic Sciences. den Haag, 1972. – PP. 14–33.
7. Fant G. Speech Sounds and Features. – Cambridge, MA, 1973.
8. Flanagan J. L. Speech Analysis, Synthesis and Perception. – Berlin, 1972.
9. Fry D. B. Acoustic Phonetics. A Course of Basic Readings. – Cambridge, MA, 1976.
10. Klatt D. Linguistic uses of segmental duration in English. Acoustic and perceptual evidence // Acoustical Society of America 59, 1976. – PP. 1208–1221.
11. Ladefoged P. Elements of Acoustic Phonetics. – Chicago, 1972.
12. Lehiste I. Readings in Acoustic Phonetics. – Cambridge, MA, 1967.
13. Lisker L., Abramson A. A cross-language study in voicing in initial stops. Acoustical measurements. – Word 20, 1964. – PP. 384–422.
14. Neppert J., Pétursson M. Elemente einer akustischen Phonetik. – Hamburg, 1986.
15. Pompino-Marschall B. Einführung in die Phonetik. – Berlin, 2009.
16. Pompino-Marschall B., Steriopolo E., Žygis M. (im Druck), Ukrainian // Journal of the International Phonetic Association.
17. Potter R. K., Kopp G. A., Green H. Visible Speech. – New York, 1947.
18. Rabiner L. R., Schafer R. W. Digital Processing of Speech Signals. Englewood Cliffs, 1978.
19. Rausch A. Untersuchungen zur Vokalartikulation im Deutschen // Beiträge zur Phonetik von Heinrich Kelz und Arsen Rausch. IPK-Forschungsberichte. – Bonn 30. – Hamburg, 1972. – PP. 35–82.
20. Rosen S., Howell P. Signals and Systems for Speech and Hearing. – London, 1991.
21. Stevens K.N. Acoustic Phonetics. Cambridge MA, 1998.
22. Tillmann H. G. (mit Mansell, Ph.) Phonetik. Lautsprachliche Zeichen, Sprachsignale und lautsprachlicher Kommunikationsprozess. – Stuttgart, 1980.
23. Titze I. R. Biomechanics and distributed-mass models of vocal fold vibration // K. N. Stevens, M. Hirano [ed.]: Vocal Fold Physiology. Proc. of the Vocal Fold Physiology Conference, Kurume. – Jan. –15–19. – Tokyo, 1981. – PP. 245–270.

\*\*\*\*\*

*Rudkivskyy, Oleksandr*  
*Nationale Linguistische Universität Kyjiw*

### **DAS PROBLEM VON *TERTIUM COMPARATIONIS* BEIM VERGLEICH DER PHONEMSYSTEME DER GERMANISCHEN UND DER SLAWISCHEN SPRACHEN**

*У статті розглянуто проблему основи зіставлення та метамови при аналізі реалізації голосних і приголосних фонем сучасної німецької, англійської, нідерландської, української, російської та польської мов та схарактеризовано вагомість фонологічного аналізу з позиції*