

Moderne Methoden der Signalanalyse
Abschlußbericht

**Grundfrequenzanalyse
musikalischer Signale**

Edgar Berdahl
Juan José Burred

Inhalt:

1) Einleitung	2
2) Technischer Hintergrund	
2.1) Signalaufbereitung.....	3
2.2) Suche nach Maxima.....	4
2.3) Maximum-Likelihood Schätzwert.....	6
2.4) Statistisches Filter.....	7
3) Ergebnisse	8
3.2) Ausblick	9
4) Bedienung	9
5) Literatur	10

1 Einleitung

Im vorliegenden Bericht wird ein Algorithmus für die Analyse der Grundfrequenz eines Klangsignals, und seine Implementierung auf MATLAB, erläutert. Die Methode die wir gewählt haben (Maximum Likelihood im Frequenzbereich) ist für die Verarbeitung musikalischer Signale besonders geeignet. Das Prinzip der Grundfrequenzanalyse wird in der Abb. 1 gezeigt [1]: das Audiosignal, das der Spieler beim Interpretieren des Musikabschnittes erzeugt, wird vom System verarbeitet, wobei die gezeigte Grafik, die den zeitlichen Verlauf der Tonhöhe darstellt, entsteht.

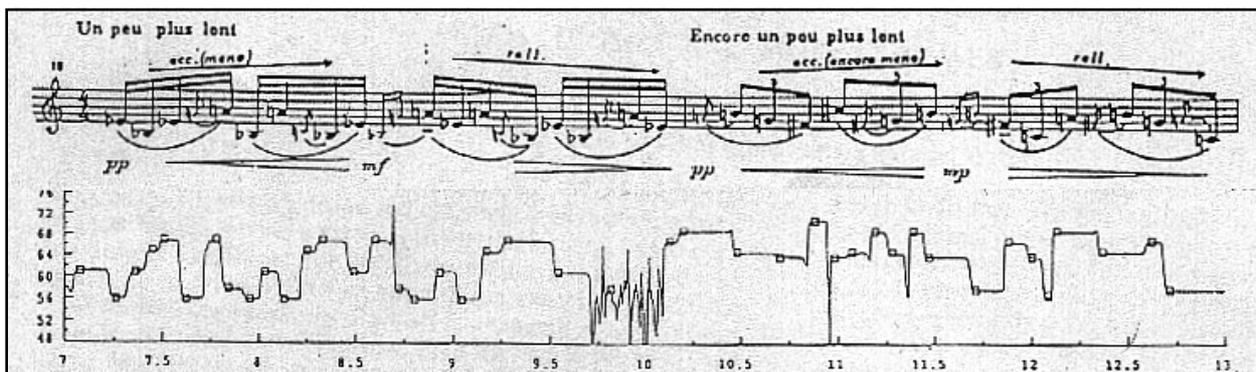


Abb. 1: Prinzip der Grundfrequenzanalyse (aus [1]).

Die automatische Bestimmung der Tonhöhe findet zahlreiche Anwendungen im Gebiet der Musiksignalverarbeitung. Sie wird z.B. in den verschiedensten Audioverarbeitungsprogrammen verwendet (bei Pitch-Shifting, Auto-Tuning, Klangeffekten...), oder bei der automatischen Transkription in Noten oder in MIDI-Daten. Bei den neulich entwickelten Systemen zur Inhaltsbeschreibung (z.B. bei MPEG-7) spielt sie eine besonders wichtige Rolle. Eine weitere, mögliche Anwendung ist die sogenannte tonhöhesynchrone Analyse. Es handelt sich um eine Kurzzeitspektralanalyse, bei der die Länge der gefensterten Signalblöcke ständig an die Tonhöhe angepasst wird, so dass in jedem Block immer die selbe Anzahl von Perioden analysiert wird. So erreicht man ein optimales Verhältnis zwischen zeitlichen und Frequenzauflösung [2].

Zahlreiche Methoden sind für die Bestimmung der Grundfrequenz vorgeschlagen oder entwickelt worden [3], doch zur Zeit existiert keine, die sowohl recheneffizient als auch beliebig zuverlässig bei einem breiten Frequenzumfang ist. Zu den Methoden die auf eine Analyse im Zeitbereich basieren, gehören z.B. die Autokorrelation-Methoden [4], die zeitliche Maximum-Likelihood Methoden, und die Verwendung von adaptiven Filtern [5]. Auf eine Analyse im Frequenzbereich basieren die Cepstrum-Methoden [6], die Maximum-Likelihood im Frequenzbereich Methoden ([1],[7],[8]), die Kombinationen von Fouriertransformationen [9] und das sog. Harmonic Product Spektrum [10]. Andere vorgeschlagene Techniken sind z.B. die Simulation der menschlichen Wahrnehmung mittels neuronaler Netzen oder die Verwendung von Wavelets [11].

Für unser Projekt haben wir uns für die Methode der Maximum-Likelihood (ML) im Frequenzbereich entschlossen, da sie für musikalische Signalen besonders geeignet ist. Außerdem ist eine recheneffiziente Implementierung möglich, wie z.B. beim im Max/MSP vorhandenen Objekt `fiddle~` [12]. Ein weiterer Vorteil ist ihre Fähigkeit, mit Rauschen und nichtharmonischen Komponenten umzugehen.

2 Technischer Hintergrund

Der ganze Algorithmus ist im folgenden Schema dargestellt (Abb. 2).

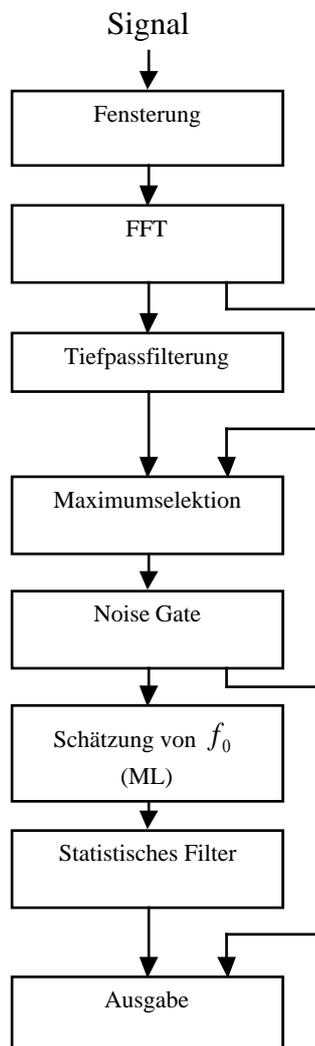


Abb. 2: Projektschema.

Im Folgenden werden wir die im Schema gezeigten Teilprozessen des Programms erläutern.

2.1 Signalaufbereitung

Fensterung und FFT

Zunächst wird das Signal für die Analyse in geeigneter Weise aufbereitet. Das Signal (in unserem Fall eine .wav oder .au Datei) wird zuerst auf einen einzigen Kanal gemischt, falls es mehrkanalig ist. Dann wird es auf 11,025 kHz umgetastet, falls die ursprüngliche Abtastrate anders ist. Man führt anschließend eine Zerlegung des Signals in überlappende Blöcke der Länge 8192 Abtastwerte durch. Wie oft eine Kurzzeitanalyse durchgeführt wird, d.h. in welcher Maß sich die Blöcke überlappen, ist vom Benutzer mittels der Benutzeroberfläche direkt einstellbar (siehe unten).

Um eine höhere Frequenzauflösung zu erreichen, wird die zweite Hälfte jedes Blocks durch Nullen ersetzt (Zero-Padding). Anschließend wird die erste Hälfte mit einem Hanning-Fenster multipliziert, und vom ganzen Block die 8192-Punkt-FFT berechnet.

Tiefpassfilterung

Ein optionales Tiefpassfilter, das die tiefen Frequenzen verstärkt und den Rest mit gleicher Amplitude durchlässt, kann auf das Spektrum angewandt werden. Ihr Frequenzgang wird in Abb. 3 gezeigt. Dieses Filter soll verwendet werden, wenn die Grundfrequenz des zu analysierenden Musikinstruments in tiefen Frequenzbereichen viel weniger Energie als die erste Partielle besitzt (z.B. beim Klavier). Damit erreicht man, dass mehr tiefe Noten (zwischen den ersten und nullten Oktaven) richtig erkannt werden.

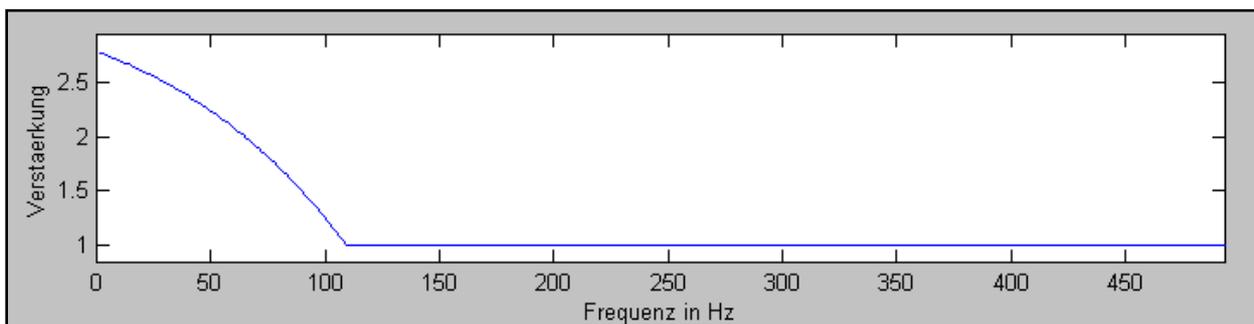


Abb. 3: Tiefpassfilter.

2.2 Suche nach Maxima

Nachdem man ein Fenster des Eingangssignals in den Frequenzbereich überführt hat, muss man entscheiden, was die Grundfrequenz dieses Fensters ist. Dazu benutzen wir, wie oben erwähnt, eine Maximum-Likelihood-Funktion, die später (siehe Kapitel 2.3) beschrieben wird. Die Eingabe zu dieser Funktion ist eine Liste der stärkeren Maxima im Frequenzbereich. Die Bestimmung dieser Liste wird im Folgenden erklärt.

Wenn man sich selber das Spektrum anschaut, ist es nicht schwierig, die starken Maxima zu erkennen; es ist jedoch nicht immer leicht für einen Computer. Wie man in der Abb. 4 sieht [13], gibt es stärkere und schwächere Maxima. Die stärkeren werden mit „actual peaks“ im Bild bezeichnet. Aus der mathematischen Analyse ist es bekannt, dass an der Stelle eines Maximums oder Minimums die Ableitung gleich Null ist. Das Spektrum ist aber im Computer natürlich nicht kontinuierlich, sondern quantisiert. Das heißt, der Nullpunkt würde sich wahrscheinlich zwischen den Abtastwerten befinden. Deswegen schaut man sich die Abtastwerten im Frequenzbereich von links, den tiefen Frequenzen entsprechend, bis rechts, den höheren Frequenzen entsprechend, an und sucht nach einer Stelle, wo sich die Ableitung von positiv nach negativ ändert.

Wenn man aber einfach diesen Vorzeichenwechsel der Ableitung des Spektrums als Kriterium benutzen würde, würde man manchmal sehr viele Kandidaten für die Maxima bekommen, und das wären zu viel für eine vergleichsweise schnelle Berechnung der Maximum-Likelihood-Funktion. Diese Kandidaten wären auch manchmal ziemlich schwache Maxima, die wahrscheinlich nichts mit der Grundfrequenz oder deren harmonischen Partiellen zu tun hätten.

Diese starken Maxima werden mit lila Kreisen im Spektrum in dem ersten Plot im Programm gekennzeichnet (siehe Abb. 6).

Nach diesem Schritt hat man eine Liste von manchmal sehr viele (bei einem rauschhaften Signal zum Beispiel) und manchmal weniger Maxima. Man möchte die Länge der Liste begrenzen, damit die Maximum-Likelihood-Funktion ziemlich schnell funktioniert. Deswegen wird die Liste nach der Amplitude der Maxima sortiert, falls die Länge größer als die Konstante `maximums` ist, und die kleineren werden dann entfernt. Das Sortieren ist ein Selection-Sort-Algorithmus, der durch die `max` Funktion von Matlab implementiert wird. Die gekürzte Liste wird dann an die Maximum-Likelihood-Funktion übergeben, und diese stärksten Maxima werden mit grünen Sternen im ersten Plot im Programm gezeigt.

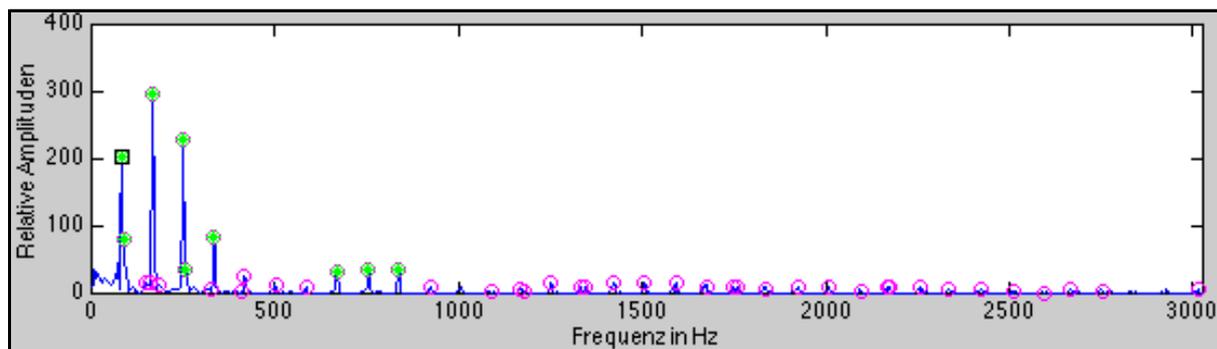


Abb. 6: Das erste Plot im Programm: ein Spektrum, wobei die stärkeren Maxima mit lila gekennzeichnet werden, die stärksten Maxima mit grünen Sternen (es gibt höchstens eine Anzahl von `maximums` von denen). Die erkannte Grundfrequenz wird mit einem schwarzen Rechteck gezeigt. Es wird unterhalb der Frequenz `lowestFreq` nicht gesucht—deshalb gibt es in diesem Fall keine Kreise unterhalb von 80 Hz.

Nur wenn mindestens ein Maximum einen bestimmten Schwellwert überschreitet wird die Grundfrequenzsuche durchgeführt, sonst wird das ganze Prozess übersprungen (Noise Gate), und es wird keine Note am Ausgang gezeigt.

2.3 Maximum-Likelihood Schätzwert

Nun hat man eine Liste spektraler Maxima, die als Kandidaten für die Grundfrequenz betrachtet werden können. Die Aufgabe des folgenden Programmblocks ist, aus den Kandidaten, das Maximum zu wählen, das am wahrscheinlichsten der Grundfrequenz entspricht. Dazu berechnet man für jeden Kandidaten (jeder Kandidat entspricht einer Frequenz f) seinen sogenannten Maximum-Likelihood Schätzwert $L(f)$, der in diesem Fall durch folgende Formel definiert werden kann:

$$L(f) = \sum_{i=1}^k a_i t_i n_i$$

wobei $1 \leq k \leq \text{maximums}$. Die erkannte Grundfrequenz f_0 ist die Frequenz des Maximums mit dem größten ML-Wert. Mathematisch betrachtet handelt es sich hier um eine Maximierung der Art:

$$f_0 = \arg \max \{L(f)\}$$

Die Parameter a_i , t_i und n_i haben folgende Bedeutung:

Amplitudenparameter a_i

Der Amplitudenparameter a_i ist von der Amplitude A_i des i -ten Maximums abhängig. Ein Maximum mit höherer Amplitude bekommt einen größeren ML-Wert als ein Maximum mit einer kleineren Amplitude und mit den gleichen t und n Parametern. Anders ausgedrückt: die Grundfrequenz ist wahrscheinlich eine der Frequenzen der größten Maxima der Liste. Wir haben verschiedene Formeln für die Berechnung dieses Parameters ausprobiert, und haben die besten Ergebnisse mit einem logarithmischen Verhältnis bekommen, laut der Formel:

$$a_i = C \log A_i$$

wobei C ein Proportionalitätsparameter ist.

Harmonizitätsparameter t_i

Für jedes untersuchte Maximum werden die restlichen Maxima der Liste, die einer Frequenz entsprechen, die höher als die des untersuchten Maximums ist, mit den genauen harmonischen Frequenzen dieses Maximums verglichen. Dieser Harmonizitätsparameter ist entscheidend, um eine richtige Grundfrequenzerkennung zu bekommen; man muss ihn dementsprechend sorgfältig modellieren. Nach dem in der Literatur vorgeschlagenen Wahrscheinlichkeitsmodell ([1],[7],[8]) haben wir dafür eine Gaußsche Funktion der Form:

$$t_i = Ce^{-\frac{d_i^2}{D}}$$

verwendet, wobei d_i der Frequenzabstand zwischen Kandidaten und Harmonischen, und C und D Proportionalitätsparameter sind.

Frequenzparameter n_i

Je tiefer die Frequenz eines Kandidatsmaximums ist, desto wahrscheinlicher entspricht er der Grundfrequenz. Dieser Verhältnis wird von uns exponentiell modelliert:

$$n_i = Ce^{-\frac{f_i}{D}}$$

wobei f_i die Frequenz des Kandidaten, und C und D Proportionalitätsfaktoren sind.

2.4 Statistisches Filter

Die 5 letzten erhaltenen Grundfrequenzwerte werden in einem Buffer gespeichert, und miteinander verglichen, wobei die Varianz berechnet wird. Nur wenn diese Varianz kleiner als eine bestimmte, in der Variable `maximumVariance` festgelegte maximale Varianz, ist, wird die neue Grundfrequenz als gültig betrachtet und zur Ausgabe geliefert. Dies dient dazu, Fehler bei den Notenübergängen zu verringern.

3 Ergebnisse

Die folgende drei Bilder (Abb 7-9) zeigen verschiedene Ergebnisse des Programms (MIDI-Nummer¹ gegen Zeit). Um die beste Ergebnisse zu bekommen muss man die spektrale Eigenschaften des analysierten Instruments bzw. Aufnahme betrachten, und die Parameter entsprechend einzustellen (siehe unten: Benutzeroberfläche).

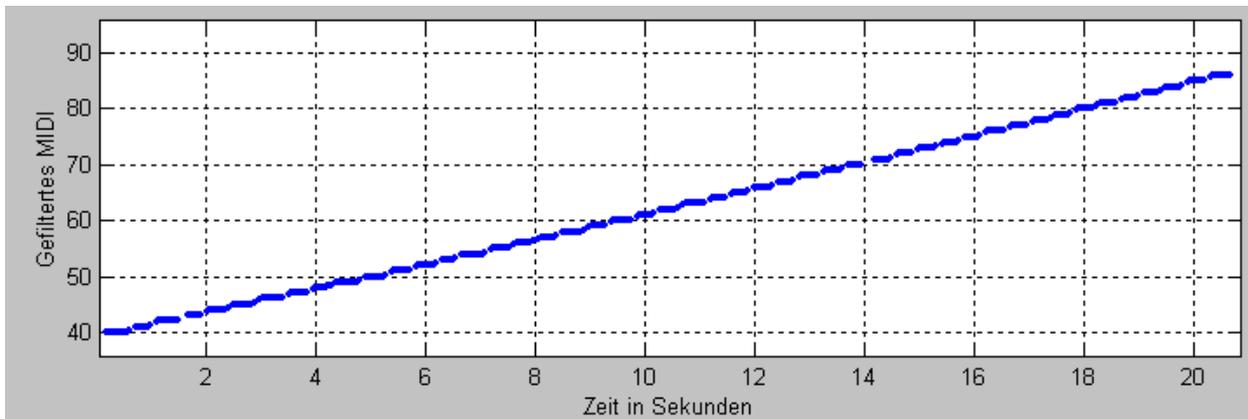


Abb. 7: Ergebnis für eine unverzerrte elektrische Gitarre (chromatische Skala)

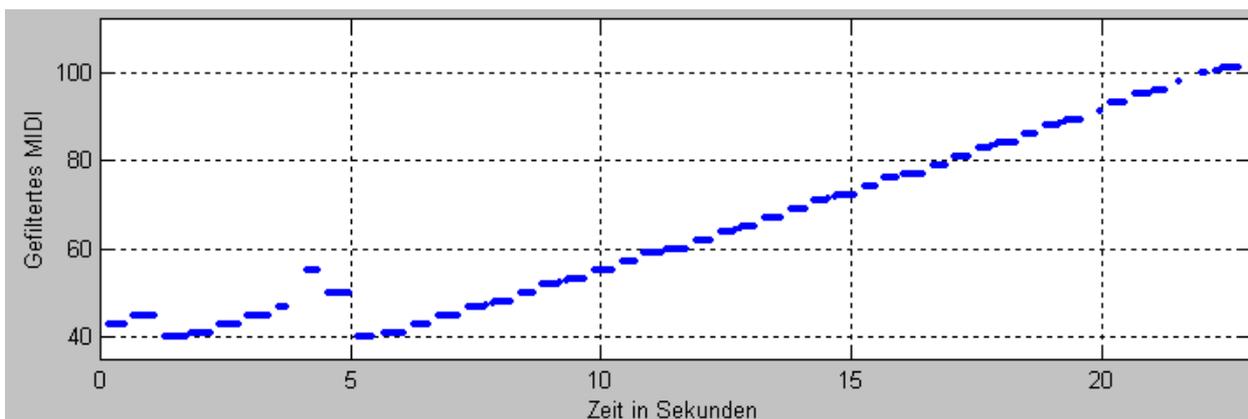


Abb. 8: Ergebnis für ein Klavier (diatonische Skala)

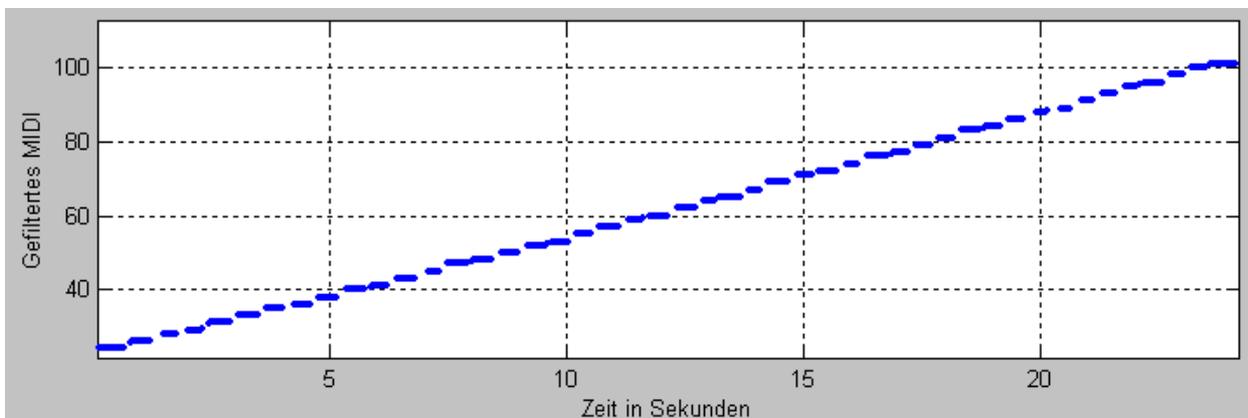


Abb. 9: Ergebnis für ein Keyboard (diatonische Skala)

¹ MIDI 40 = e1; MIDI 100 = e6

Die meisten Probleme sind bei Aufnahmen mit viel Nachhall, bei der Erkennung der tiefsten Noten und bei Signalen, in welchen die Grundfrequenz sehr kaum vorhanden ist, aufgetreten.

3.2 Ausblick

Als wir mit diesem Projekt anfangen, waren wir uns ziemlich sicher, dass die Maximum-Likelihood-Methode eine der besten war. Obwohl wir in nachhinein immer noch dieser Meinung sind, können wir jetzt verstehen, warum es so viele Methoden gibt und warum keine Methode für alle musikalische Signale ideal funktioniert. Die Maximum-Likelihood-Methode ist ein guter Ansatz und liefert zwar oft gute Ergebnisse, aber es hat sich trotzdem immer herausgestellt, dass unser Algorithmus besser funktioniert, wenn man bestimmte Vorkenntnisse über das Signal hat und ausnutzt (zum Beispiel tiefste Frequenz und zeitliche Auflösung). Deswegen können wir jetzt auch verstehen, warum es auch Methoden für die Grundfrequenzanalyse gibt, die Neuronale Netze einbeziehen, da eine bestimmte Anpassung der Einstellungen für eine Quelle sehr wichtig für den Erhalt guter Ergebnisse ist. Wir würden uns gerne in der Zukunft mit anderen Methoden der Grundfrequenzanalyse beschäftigen, oder vielleicht auch weiter mit dieser, aber wir haben uns schon bemüht einen robusten Algorithmus zu entwickeln. Um diese Methode besser zu machen, müssten wir sie wahrscheinlich mit anderen Methoden kombinieren.

4 Bedienung

Das Hauptprogramm (Analyse und Benutzeroberfläche) ist in der MATLAB-Script Datei *gfa.m* enthalten. Um das Programm zu starten muss man *gfa* im MATLAB Command Window eintippen, ohne Argumente. Es erscheint dann die Benutzeroberfläche (Abb. 10).

Ins Dateiname-Feld wird die zu analysierende Aufnahme eingegeben. Damit die Dateien gelesen werden, muss ihr Ordner im MATLAB-Path definiert sein. Wenn die Datei geladen wird, zeigt das Programm die zeitliche Wellenform. Man kann die Aufnahme wiedergeben, indem man den "Spiele"-Button anklickt. Um mit der Analyse anzufangen muss man den "Analysiere"-Button anklicken. Man kann in jedem Moment die Analyse unterbrechen durch ein zweites Drücken dieses Buttons.

Sämtliche Meldungen des Programms werden im MATLAB Command Window angezeigt.

Während der Analyse zeigt das Programm drei verschiedene Grafiken. Die oberste ist ein Teil des Spektrums jedes Frames, wobei die selektierten Maxima durch verschiedenen Farben gekennzeichnet werden (siehe auch Abb. 6). Die zweite Grafik kann entweder den Frequenzgang des Tiefpassfilters (Abb 3) oder den Grundfrequenzverlauf vor der statistischen Filterung zeigen. Dies wird durch die "Im Plot 2 anzeigen" Option bestimmt. Im zweiten Fall wird die MIDI Nummer mit der blauen Kurve, die Frequenz mit der roten Kurve, gezeigt. Die dritte Grafik zeigt das Endergebnis der Grundfrequenzanalyse (in MIDI-Nummern).

Unten rechts wird der Name der aktuellen Note angezeigt (in deutscher Nomenklatur), sowie die erkannte Grundfrequenz.

Drei Analyseparameter sind vom Benutzer einstellbar: die zeitliche Auflösung in Millisekunden (höhere Auflösung für schnelle Melodien nötig, sinnvolle Werte sind 3-50 ms), die tiefste Frequenz die zu betrachten ist (am Frequenzumfang des Instruments anpassen) und der

Schwellenwert für das Noise Gate. Außerdem kann man wählen, ob das Tiefpassfilter verwendet werden soll oder nicht.

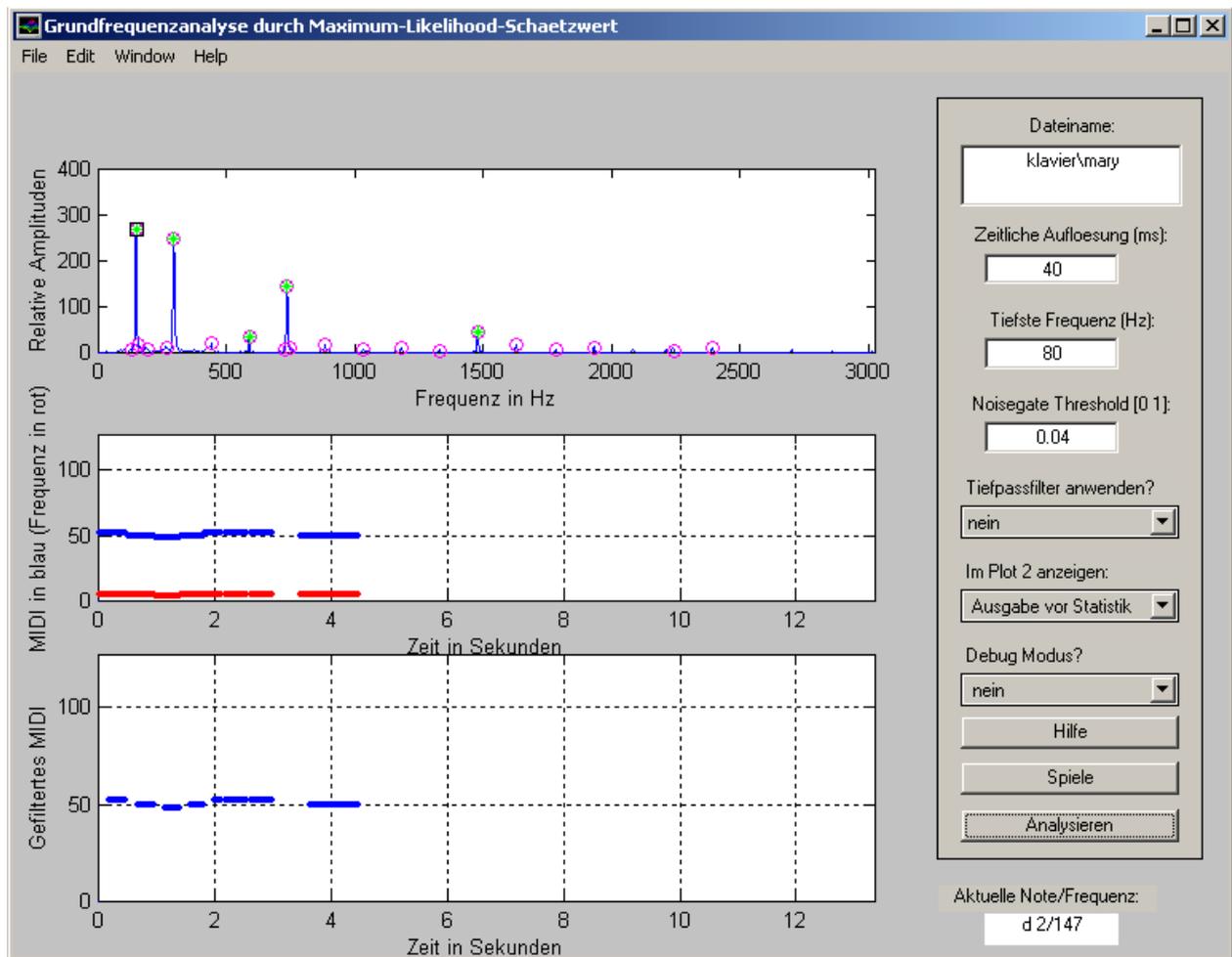


Abb. 10: Benutzeroberfläche

Das Programm ermöglicht auch ein "Debug Modus". Bei diesem Modus stoppt das Programm nach der Analyse jedes Frames und zeigt im MATLAB Command Window die Werte aller Maximum-Likelihood Parametern.

5 Literatur

- [1] DOVAL, Boris und RODET, Xavier : *Estimation of fundamental frequency of musical sound signals*. ICASSP, Band 5, S. 3657-3660. Mai 1991.
- [2] SERRA, Xavier : *Musical Sound Modeling with Sinusoids plus Noise*, veröffentlicht in *Musical Signal Processing*, Ed.: C. Roads, S. Pope, A. Picialli, G. De Poli, Swets & Zeitlinger Publishers, 1997
- [3] HESS, W. : *Pitch Determination of Speech Signals*, Springer Verlag, Berlin, 1983.
- [4] BROWN, Judith C., ZHANG, Bin : *Musical Frequency Tracking Using the Methods of Conventional and "Narrowed" Autocorrelation*. J. Acoust. Soc. Am. 89 (5), Mai 1991.

- [5] LANE, John E. : *Pitch Detection Using a Tunable IIR Filter*, Computer Music Journal, vol. 14, no. 3, 1990
- [6] NOLL, Michael A. *Cepstrum Pitch Determination*. J. Acoust. Soc. Am. 41 (2), 1967.
- [7] DOVAL, Boris und RODET, Xavier. *Fundamental frequency estimation and tracking using maximum likelihood harmonic matching and HMMs*. ICASSP, Band 1, 221-224. 1993.
- [8] STARKE, Jens-Alexander : *Grundfrequenzanalyse nach dem Wahrscheinlichkeitsmodell von Boris Doval und Xavier Rodet*. Diplomarbeit, Technische Universität Berlin, FB Elektrotechnik
- [9] MARCHAND, Sylvain : *An efficient pitch-tracking algorithm using a combination of fourier transforms*. In Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-01), Limerick, Ireland, December 6-8, 2001. Auch unter <http://www.csis.ul.ie/dafx01/proceedings/papers/marchand.pdf>
- [10] SCHROEDER, M.R. : *Period Histogram and Product Spectrum: New Methods for Fundamental Frequency Measurements*, J. Acoust. Soc. Am. 43, 829-834, 1968
- [11] JEHAN, Tristan : *Musical Signal Parameter Estimation* . Thesis, CNMAT, 1997
- [12] PUCKETTE, Miller S., APEL und ZICARELLI. *Real-time audio analysis tools for Pd and MSP*. ICMC Proceedings, 1998.
- [13] PARK, Tae Hong. *Salient Feature Extraction of Musical Instrument Signals*. Master's Thesis, Dartmouth College in Hanover, New Hampshire, USA. 2 Juni 2000.
- [14] Puckette, Miller S. *Pure Data: Recent Progress*. In Proceedings, Third Intercollege Computer Music Festival, Tokyo, pp. 1-4. Auch unter <http://man104nfs.ucsd.edu/mpuckett>
- [15] Brown, Judith C. und Miller Puckette. *A high resolution fundamental frequency determination based on phase changes of the Fourier transform*. J. Acoust. Soc. Am. 94 (2), Pt. 1, August 1993.
- [16] Yoo, Lilit und Ichiro Fujinaga. *A comparative latency study of hardware and software pitch-trackers*. <http://gigue.peabody.jhu.edu/~ich/research/icmc99/icmc99.pitch.pdf>