



Tonhöhenwahrnehmung bei Lamellophonen

Project Manager: Alois Sontacchi (sontacchi@iem.at)

Contributors : Gerd Grupe
Robert Höldrich
Gerhard Nierhaus

Tonhöhenwahrnehmung bei Lamellophonen und die Entwicklung eines computergestützten Tools für interaktive Experimente zu Stimmung und Tonsystem.

1. Einleitung

In diversen Untersuchungen [1] zu Idiophonen wird eine einheitliche und eindeutige wahrgenommene Tonhöhe aufgrund der komplexen und teilweise inharmonischen Spektren angezweifelt. Im Rahmen dieser Untersuchung sollte zunächst die Hypothese getestet werden, dass Metalllamellen (Lamellophone, siehe Abb.1) überwiegend Spektraltonhöhen produzieren. Des Weiteren sollten auch ggf. die Existenz von Tonhöhenverschiebungen geprüft werden. Vorerst werden dazu Untersuchungen mit „westlichen“ Versuchspersonen durchgeführt. In einem zweiten Schritt soll ein computerunterstütztes System entwickelt werden, das interaktive Experimente zur Stimmung der Lamellophone erlaubt.



Abbildung 1: Lamellophon

2. Aufgabenstellungen der Untersuchungen und Vorgangsweise

Auflistung der Arbeitsschritte und Überblick der Untersuchung:

- Aufnahme von Lamellophonen (*Dambozoko*, *Mhlanga*, *Gadhanga*), wobei Instrumente voraussichtlich unterschiedliche Stimmungen besitzen.
- Analyse der aufgenommenen Lamellenklänge basierend auf dem gehörgerechten Modell nach Terhardt (1982, [2]).
- Analyse der aufgenommenen Lamellenklänge durch Korrelogramm und Periodogramm.
- Durchführung eines präliminären Hörversuches zur Tonhöhenerkennung
- Auswertung der Hörversuchsdaten
- Realisierung eines Analyse-Resynthese-Tools
- Interaktive Experimente
- Fazit

2.1 Ergebnisse des Modells nach Terhardt

Terhardt beschreibt in [2] ein Analysemodell zu Detektierung und Bestimmung der am wahrscheinlichsten wahrgenommenen Tonhöhe basierend auf der spektralen Analyse. Das Eingangssignal wird durch eine Kurzzeit-Fourier-Transformation in den Spektralbereich gebracht und bezüglich tonaler (deterministischer) und geräuschartiger (stochastischer) Komponenten untersucht. Für all jene tonale Komponenten die über der Ruhehörschwelle liegen werden die gegenseitigen Maskierungsschwellen bestimmt. Durch Simultanmaskierung verjüngt sich das Set von möglichen Kandidaten. Der Pegel der verbleibenden Tonhöhenkandidaten und darüber hinaus noch deren Lage im Spektrum (Berücksichtigung der spektralen Dominanz: 500 - 2500 Hz) liefern Beiträge für das Gewichtungskriterium (siehe unten). Des weiteren sind jene verbleibende Komponenten Anwärter für die Berechnung von virtuellen Tonhöhen. Die Berechnung erfolgt nach Bestimmung der subharmonischen Koinzidenz.

Das gehörgerechte Analysemodell nach [2] wurde in Matlab implementiert und anhand der veröffentlichten Ergebnisse [3] auf die Reproduktionsgenauigkeit überprüft. Das Modell nach Terhardt liefert für die zu erwartenden wahrgenommenen Tonhöhen einen Gewichtungsfaktor der dadurch eine Reihung ermöglicht. Durch Skriptdateien wurden die einzelnen Klangfiles analysiert und die Ergebnisse nach der Gewichtsreihung und mit den Vermerk, ob es sich um eine virtuelle oder spektrale Tonhöhe handelt, in Listen gespeichert.

2.2 Ergebnisse mit Hilfe des Korrelogramms und Periodigramms

Dieses Analysesystem basiert grundsätzlich auf der Korrelation im Zeitbereich, wobei die Modellierung des peripheren Gehörs in der Signalverarbeitung berücksichtigt wird. Als Ausgangspunkt dient das Signalmodell aus [4].

Das Eingangssignal wird durch eine Filterbank - entspricht der Frequenzanalyse entlang der Basiliarmembran - in Frequenzkanäle unterteilt. Die jeweiligen Signal der Frequenzkanäle durchlaufen eine Halbwellenleichte - vereinfachtes Haarzellenmodell - und werden anschließend durch eine Autokorrelation hinsichtlich ihrer Periodizität untersucht. Dadurch erhält man vorläufig für einen Zeitausschnitt ein Korrelogramm d.h. Korrelationsfunktionen für jeden Frequenzkanal. Eine zeitliche Abfolge dieser Korrelogramme für sequentielle Zeitausschnitte aus dem Originalsignal (siehe Abb. 2) wird auch ‚perceptual movie‘ genannt. Man erkennt in Abhängigkeit vom Zeitpunkt in den einzelnen Bändern die dominante Periodizität.

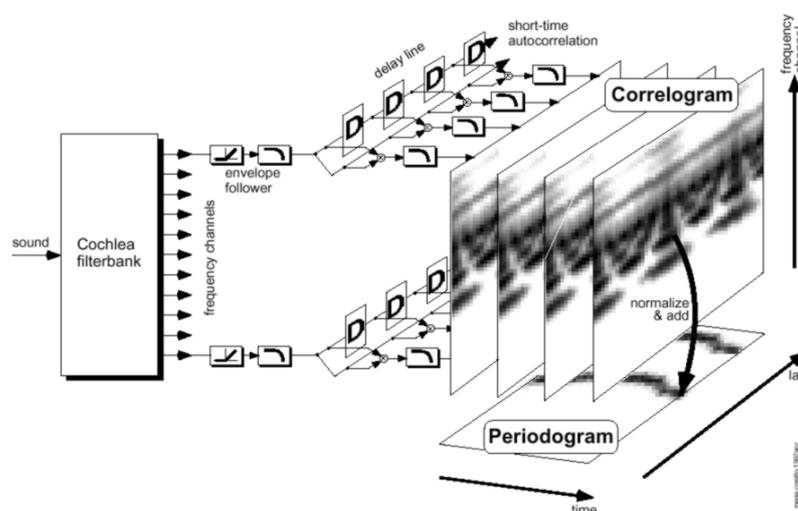


Abbildung 2: Zeitbasierendes Modell (Korrelogramm & Periodogramm) der peripheren Verarbeitung mit vereinfachtem Haarzellenmodell.

Durch eine Integration der Autokorrelationsfunktion über alle Frequenzkanäle hinweg liefert, für unterschiedliche Zeitverschiebungen (Lags) der Frequenzkanalsignale, die jeweilig stärkste Periodizität. Eine zeitliche Verfolgung dieses Prozesses führt dann zum Periodigramm (siehe Abb. 2). Durch eine signalangepasste zeitliche Integration des Periodigramms und eine zusätzliche Berücksichtigung der spektralen Dominanz führt nach Bestimmung der relevantesten Periodizität – stärksten Frequenzkomponente zur voraussichtlichen grundlegenden Tonhöhe.

Für die optimale Bestimmung der Grundtonhöhe ist eine Anpassung der folgenden Parameter an die Signaleigenschaften unumgänglich. Dies sind das Außen- und Mittelohr-Gewichtungsfiler (AMO), die Gammatonfilterbank (Cochlea), der Grad der Nichtlinearität des Haarzellenmodells, das Glättungsfiler und die Länge Autokorrelation, sowie die zeitliche Integration des Periodigramms.

2.3 Beschreibung des informellen Hörversuches

Der Hörversuch wurde vorerst für eine kleine Teilnehmeranzahl von 9 Probanden beschränkt. Dieser besteht aus 8 Aufgaben. Jede Aufgabe stellt einen AB Versuch dar. Die Aufgabenstellung an die Probanden besteht darin die Frequenz eines Sinustones – Klang B – identisch der wahrgenommenen Frequenz des Klanges A mit Hilfe von endlosdrehbaren und unbeschrifteten Drehreglern (3 Regler mit unterschiedlicher Quantisierung: Halbtonintervall, +/-20% um Startfrequenz, +/-2% um Startfrequenz) einzustellen (vgl. dazu Kap.11, S.306 ff in [5]). Der Klang A wird von einer Versuchsaufgabe zur nächsten variiert. Der Versuch wird in 3 Sitzungen á 60min aufgeteilt. In der ersten Einheit wird das Hörprofil der Probanden bestimmt. In der zweiten Einheiten werden einzelne ausgewählte Lamellen untersucht (Reliabilität und Konsistenz). In der dritten werden unterschiedliche Instrumente (*Dambatsoko, Mhlanga, Gadhanga*) durchgestimmt.

2.4 Erste Ergebnisse des Hörversuchs

Aus einer ersten informellen Sichtung der Probandendaten kann bereits folgendes berichtet werden: Die Probanden neigen dazu, wie vermutet, spektralen Tonhöhen den Vorzug bzgl. der wahrgenommenen Tonhöhe zu geben. Es bilden sich musikalische Tonkategorien aus. Einzelne geübte Hörer zeigen jedoch eine beachtliche hohe Genauigkeit bei der reproduzierbaren Tonhöhenbestimmung. Des weiteren ist der Unterschied bzgl. Sicherheit und Konsistenz der wahrgenommenen Tonhöhe bei Tonfolgen (holistisch) gegenüber Einzelklängen (analytisch) erkennbar. Da bei der Auswertung der Hörversuchsdaten jedoch nur 3 Probanden für hochqualitative (d.h. genaue, konsistente und reliable) Antworten extrahiert werden konnten sind alle statistischen Angaben (Tonhöhenmittelwerte und Varianzen) nur bedingt brauchbar - zulässig.

2.5 Analyse und Resynthese der Lamellophonklänge

Der Klang einer Lamelle kann als eine Summe von exponentiell gedämpften Sinusschwingungen aufgefasst werden (siehe Gl. 1). Daher liegt es nahe die Analyse der Lamellophon Klänge mit dem Steiglitz-McBride Ansatz [6] zu lösen.

$$s(n) \approx \sum_{k=1}^K a_k(n) \cdot e^{-\alpha \cdot n} \cdot \sin(2\pi \cdot f_k \cdot n + \phi_k) \quad \text{Gl.1}$$

Der Ansatz von Steiglitz-McBride ist ein iteratives Verfahren, basierend auf der Prony-Methode, welches einen Signalausschnitt (Impulsantwort) durch eine Summe von Pole (Frequenz f_k und Resonanzgüte α_k) innerhalb des Einheitskreises (bzw. entlang der Frequenzachse) modelliert. Aus den Residuen der Übertragungsfunktion wird die Amplitude a_k und Startwinkel ϕ_k der einzelnen gedämpften Sinusschwingungen bestimmt. Durch 32 frequenzselektive Subbändern konnte die Analyse verbessert werden. Dadurch ist die Verteilung der Pole pro Frequenzband steuerbar (siehe exakte Rekonstruktion im tieffrequenten Bereich in Abb. 3) auch bei geringer Gesamtanzahl von Analyseparameter.

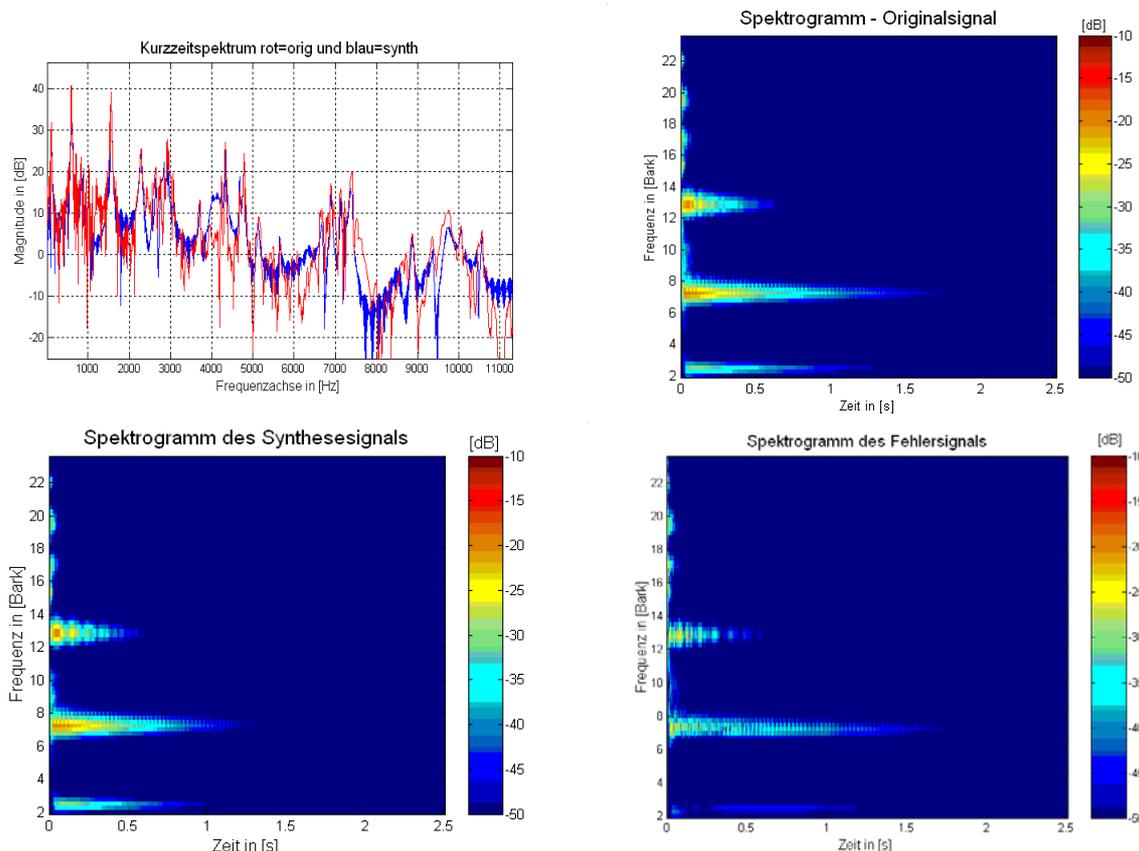


Abbildung 3: Ergebnisse des Analyse-Resynthese-Tools in Matlab.

Oben Links: Spektrale Schätzung des Originalsignals mit 48 Polstellen.

Die Spektren (rot = Original, und blau = Synthese) zeigen jeweils den klangbestimmenden Signalausschnitt zu Beginn eines Lamellophonklanges.

Oben Rechts: Spektrogramm des originalen Klanges.

Unten Links: Spektrogramm des re-synthetisierten Klanges.

Unten Rechts: Spektrogramm des Fehlersignals (Energie der Abweichungen)

Die Implementierung der Resynthese erfolgt mit pure data (PD) einer grafischen Computermusik Programmiersprache. Anhand einer grafischen Bedienoberfläche können einzelne Signalparameter für die Resynthese selektiv modifiziert werden. Dabei werden die Presets eine Oszillatorbank (gedämpfte Sinusschwingungen, Variablen siehe Gl. 1) durch die Parameter aus der Analysestufe (Matlab Tool) festgelegt. Für die Interaktivität mit den einstellbaren Klängen werden die vier prominentesten Teilklangkomponenten erstgereiht. Bei den Feldversuchen steht somit vorerst nur ein weiter reduzierter (in Echtzeit) variabler Parametersatz (4 Frequenzen und 4 Amplituden) zur Verfügung.

3 Status Quo

Die Ergebnisse aus der Korrelogramm basierten Analyse stimmen weitgehend mit den Hörversuchsergebnissen gut überein (Übereinstimmung für 70% innerhalb +/- 15 cent). Für die tiefen Register ergeben sich jedoch z. T. noch beachtliche Abweichungen (d.h. größer als ein Viertelton, beinahe Halbton). Das Terhardt Modell liefert zwar im Ergebnisset der möglichen Tonhöhenkandidaten richtige Frequenzwerte, jedoch erzeugt der Algorithmus nicht die erforderlichen Gewichte für die entsprechende Erstreuung.

Das physikalisch motivierte Signalmodell mit exponentiell gedämpften Sinusschwingungen liefert brauchbare Ergebnisse. Ein erster Prototyp für interaktive Experimente in Feldstudien wurde realisiert. Die intuitive Bedienbarkeit und Systemzuverlässigkeit muss noch an die gestellten Anforderungen angepasst werden.

Die Verifizierung der Hypothese bzgl. einer eindeutigen Tonhöhe ist informell für westliche Hörer (wenn auch statistisch nicht abgesichert) gegeben, jedoch bedingt durch die unsichere (und z. T. noch zu große) Varianz ist auf ein Tonsystem derzeit nicht schließbar.

Literatur:

- [1] Schneider, Beurmann, zahlreiche Veröffentlichungen von 1991 bis 1996.
- [2] Terhardt, E. Stoll, G., Seewann M.: „Algorithm for extraction of pitch and pitch salience from complex tonal signals“, J. Acoust. Soc. Am. 71(3), March 1982.
- [3] Terhardt, E. Stoll, G., Seewann M.: „Pitch of complex signals according to virtual-pitch theory: Tests, examples, and predictions“, J. Acoust. Soc. Am. 71(3), March 1982.
- [4] Ellis, D.: “The Welft: Auditory Scene Analysis of periodic sounds“, ICASSP’97, Poster.
- [5] Terhardt, E.: “Akustische Kommunikation“, Springer Verlag Berlin Heidelberg New York, ISBN 3-540-63408-8, 1998.
- [6] Lawrence M.S. (ed.), “Digital Spectral Analysis with Applications“, Prentice Hall Inc., ISBN: 0-13-214149-3 025, 1987.