

## Vom Bild zum Ton - Die Wiedergewinnung historischer Tondokumente durch Bildverarbeitung

Tim Wöhrle

Arbeitsgruppe Bildverarbeitung

GFaI e.V. Berlin

Rudower Chaussee 5, Geb. 13.7

12484 Berlin

Telefon 030/6392-1643 Telefax 030/6392-1602

eMail: woehrle@gfai.de, Internet: <http://www.gfai.de/projekte/spubito>

### Zusammenfassung

Der Edison-Phonograph ist das erste tonaufzeichnende Verfahren, mit dem Anfang dieses Jahrhunderts große Bestände an Tondokumenten auf Wachswalzen gewonnen wurden. Eine umfangreiche Sammlung von Abgüssen dieser Wachswalzen befindet sich im Berliner Völkerkundemuseum. Der Inhalt dieser sog. Galvanos soll möglichst berührungslos abgespielt werden. Zu diesem Zweck entwickelt die GFaI ein bildverarbeitungsgestütztes System, das mit Methoden der Bildanalyse und der 3D-Rekonstruktion die Höhe der Tonspuren vermisst. Aus dieser Spürhöhe werden die akustischen Signale rekonstruiert, und die wiedergewonnenen Klänge können direkt auf moderne Tonträger überspielt werden.

### Einleitung

Das Ende des vergangenen Jahrhunderts wurde unter anderem durch einen raschen Fortschritt auf den Gebieten der Naturwissenschaft und Technik in den westlichen Industrienationen gekennzeichnet. Zahlreiche Erfindungen wie die Telegraphie, das Automobil und das Telefon prägten die Gesellschaft. Im Zuge dieser Innovationen entwickelte der Erfinder Thomas Edison in Großbritannien das erste Aufnahme- und Abspielgerät für akustische Signale, das in der Lage war, Klänge aus der natürlichen Umwelt, also Sprache, Musik oder Gesang, aufzunehmen und mehrfach wieder abspielen zu können. Edison entwickelte das Gerät innerhalb kurzer Zeit zur Serienreife und vermarktete es unter dem

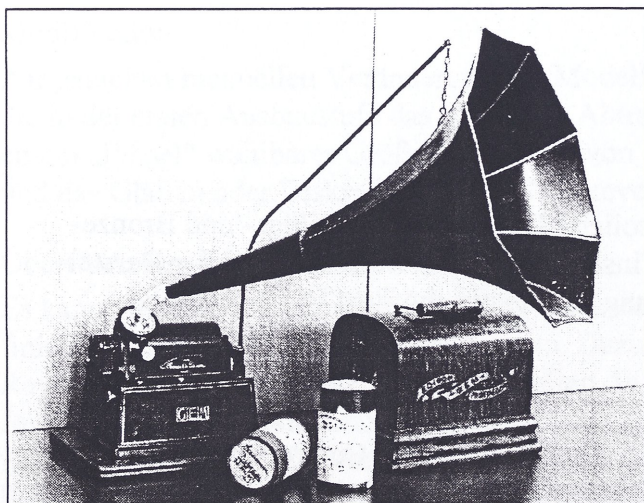


Abb. 1: Edison-Phonograph <sup>1</sup>

Namen Edison-Phonograph (Abb. 1). Der Erfolg und das öffentliche Aufsehen waren zunächst beachtlich, aber der wirtschaftliche Erfolg ließ nach kurzer Zeit nach [EXNER97], da das Verfahren noch nicht ausgereift war und nur eine unbefriedigende Tonqualität erlaubte. 1886 brachte Edison ein deutlich verbessertes System auf den Markt, bei dem erstmals auch wachsbeschichtete Walzen als Tonträgermaterial verwendet wurden (die Phonographen der ersten Generation verwendeten mit Zinnfolie bespannte Spindeln).

Das System erwies sich als wirtschaftlicher Erfolg, wurde noch mehrfach weiterentwickelt und verbessert (allerdings später nicht mehr von Edison selbst, der sich anderen Erfindungen wie der Glühlampe zuwandte) und befand sich noch bis in die 30er im Einsatz. Außer von Edison wurde es von einer Reihe von Konkurrenzunternehmen vertrieben. In dieser Spanne von 4 Jahrzehnten entstand weltweit ein großer Fundus von

<sup>1</sup> Abbildung mit freundlicher Genehmigung des Museums für Völkerkunde, Abt. Musikethnologie



Aufnahmen sehr unterschiedlicher Intention und Qualität. Neben sehr vielen Aufnahmen mit zeitgenössischer Unterhaltungsmusik entstanden auch rare oder musikwissenschaftlich interessante Aufnahmen. Die Rekonstruktion dieser Aufnahmen mit Hilfe von Bildanalyse und -bearbeitung ist das Ziel des Projektes SpuBiTo, das im folgenden vorgestellt werden soll.

### Zum Aufbau der Edisonwalzen

Der Edison-Phonograph ist heutzutage weit weniger bekannt als seine kurze Zeit später entwickelte Konkurrenz, das Grammophon. Einer der Gründe ist sicherlich in der unterschiedlichen technischen Realisierung der beiden Systeme zu suchen. Bei Grammophonen wurde die Toninformation durch horizontale Modulation auf Scheiben aufgebracht, eine Technik, die sich über die Schellackplatten bis zu den Langspielplatten durchsetzte, bis sie schließlich durch digitale Codierung der Compact Disks abgelöst wurde. Insofern kann man das Grammophon durchaus als „Großvater“ der modernen Schallplatte bezeichnen.

Beim Edison-Phonographen dagegen werden die Töne durch vertikale Modulation auf walzenförmige Tonträger aufgebracht. Diese innen hohlen Walzen werden auf einer angetriebenen Spindel bewegt, wobei die Bewegung der Nadel durch einen Vorschub mit konstanter Geschwindigkeit entlang der Rotationsachse der Walzen erfolgt. Da die Tonspuren mit einer geeigneten Nadel geschnitten werden, muß ein entsprechend weiches Material zu ihrer Herstellung verwendet werden (zunächst verschiedene Wachse, später auch „Blue Amberol“, ein Cellulidmaterial). Die Verwendung dieser Materialien bedingt eine begrenzte Abspielbarkeit der Walzen, bis die Abspielnadel das Wachs soweit abgetragen hat, daß die Klangqualität erheblich nachläßt.

Vor allem dieser Nachteil sowie die Tatsache, daß sich Edison-Walzen im Gegensatz zu Grammophon-Platten nicht einfach und in großem Maße kopieren ließen, führten schließlich in den 20er und 30er Jahren zur Durchsetzung der Grammophon-Technologie.

In folgender Tabelle sind die Eigenschaften von Phonograph und Grammophon gegenübergestellt.

Edison-Phonograph	Grammophon
Aufnahme und Abspielen mit einem Gerät möglich	Nur Abspielen möglich; Aufnahme erfordert ein anderes (und viel komplizierteres) Gerät
Mehrfachbespielen und „Löschen“ (durch Abtragen der Tonspuren) von Walzen möglich	Nur einmaliges Bespielen möglich
Einführung 1877 (dadurch 10-jähriger Entwicklungsvorsprung)	Einführung 1887
Relativ geringe Klangqualität	Höhere Klangqualität durch geringeren Auflagedruck
Empfindliches und sich abnutzendes Wachsmaterial durch Schneidverfahren	Robustes Material (Schellack) durch Gußverfahren
Zunächst ganz fehlendes und später unbefriedigendes Kopierverfahren	Effektives Kopierverfahren und damit Möglichkeit der Massenverbreitung
Platzaufwendiges Format der Walzen	Platzsparendes Format der Platten

Tab.1: Gegenüberstellung Edison-Phonograph - Grammophon

### Das Berliner Phonogramm-Archiv

Unter dem Einfluß der neuen Technologie der Tonaufnahme und -archivierung wurde Anfang des 20. Jahrhunderts das Berliner Phonogramm-Archiv gegründet, mit dem Ziel, ein Archiv von musikethnologisch interessanten Tondokumenten mit Musik von Völkern aus aller Welt zusammenzustellen. In den folgenden Jahren bat man gezielt Wissenschaftler, die eine Exkursion zu anderen Kontinenten planten, um Kooperation, und gab ihnen für ihre Reisen einen Phonographen sowie unbespielte Walzen mit.



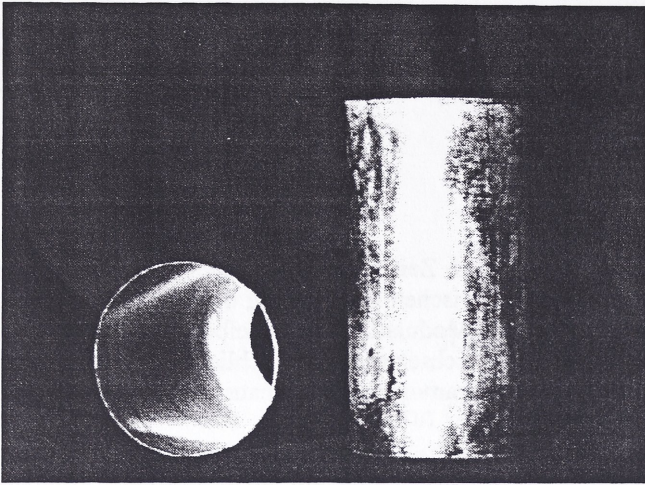


Abb. 2: Ansicht zweier Galvanos

Ursprungsort, der Musikethnologischen Abteilung des Völkerkundemuseums Berlin, die aus dem Phonogrammarchiv hervorgegangen war.

Die originalen Wachswalzen haben die lange Odyssee und die unsachgemäße Lagerung gar nicht oder in zum Teil schlechter Qualität überstanden, so daß sich der Bestand heute zum größten Teil aus Galvanos zusammensetzt. Die Sammlung umfaßt laut [EXNER97] heute:

- 2.747 originale Wachswalzen
- ca. 14.000 Galvanos
- ca. 13.000 Kopien aus Hartwachs

Wegen des schlechten Zustandes der Hartwachskopien erhofft man sich nun eine Rekonstruktion der historischen Tonaufnahmen direkt aus den Galvanos, die die lange Zeit aufgrund ihrer Materialeigenschaften deutlich besser überstanden haben. Ein erneutes Abgießen wird von Fachleuten wegen der Einmaligkeit der Walzen als zu riskant erachtet, und man strebt nach einer möglichst schonenden, berührungslosen Lösung. Im Rahmen des Projektes SpuBiTo entwickelt die GFaI ein berührungsloses, optisch abtastendes Verfahren, das auf Methoden der Bildverarbeitung und -analyse basiert. Das System besteht aus einem Tubus mit angeschlossener Kamera, der über eine ansteuerbare Achse in den Galvano gefahren werden kann und Bilder von der Innenseite ermöglicht. Der Galvano wird auf einem Mehrachsensystem gedreht und ggf. verfahren und gekippt, um vorhandene Geometriefehler auszugleichen.

## Die Vermessung der Tonspuren

Die genaue und robuste Vermessung des Höhenprofils der Tonspuren ist entscheidend für die Qualität der gewonnenen Klänge und stellt somit den zentralen Part von SpuBiTo dar. Aus den abgetasteten Spürhöhen ergibt sich durch direkte Umrechnung die Toninformation. Die Höhenabtastung durch bildanalytische Verfahren entpricht mit ihren Ergebnissen der Wirkungsweise der Abtastnadel in konventionellen Abspielgeräten und bedurfte deshalb einer sorgfältigen Planung und Implementierung. Qualitative Mängel, die an dieser Stelle entstehen, können nachträglich nicht mehr oder nur unvollständig korrigiert werden.

Abb. 3 zeigt ein Bild von mehreren Tonspuren in Draufsicht. Die Tonspuren, die bei den originalen Wachswalzen die Form von Rillen aufweisen, haben bei den Galvanos das negative Erscheinungsbild, sind also erhaben. Ihre Spürhöhe wird durch die schwingende Aufnahmenadel gebildet und macht die Amplitude und damit die Lautstärke eines Tons aus. Je lauter ein Ton in der Aufnahme, desto höher ist das Auf und Ab der Spuren.

Die Tonhöhe (Frequenz) eines Tons ergibt sich aus den Abständen der „Kämme“ einer Tonspur zueinander; je dichter also Berg und Tal beisammen sind, desto schneller schwingt die Nadel und desto höher erscheint einem Zuhörer der Ton. Abb. 3 zeigt Spuren verschiedener Frequenzen. Im Falle eines reinen Klanges (z.B. Kammerton A, 440 Hz) haben die Spürhöhen einen sinusförmigen Verlauf, was jedoch in der Praxis nur selten vorkommt.

Auf diese Weise kam in den folgenden Jahren eine umfangreiche Sammlung von Tondokumenten zustande [ZIEGLER95]. Da man sich der begrenzten Abspielbarkeit der Wachswalzen, die jede für sich ein nicht wiederbeschaffbares Unikat darstellt, bewußt war, wurde ein Kopierverfahren für die Originale entwickelt. Durch Galvanisierung wurde von jeder Walze ein Kupfernegativ (Galvano) angefertigt, aus dem dann durch Abgußverfahren neue Wachskopien gezogen werden können. Bei dem Galvano befinden sich die Abdrücke der Tonspuren auf der Innenseite (Abb. 2), wodurch die Galvanos selbst mit herkömmlichen Verfahren nicht abspielbar sind, sondern erneut abgegossen werden müssen.

Nach dem 2. Weltkrieg wurde der größte Teil des Phonogrammarchives (90% des Bestandes) von der Sowjetunion eingezogen und nach Leningrad geschafft. 1960 wurde der Bestand an die DDR übergeben und gelangte erst nach der Wiedervereinigung 1991 zu seinen



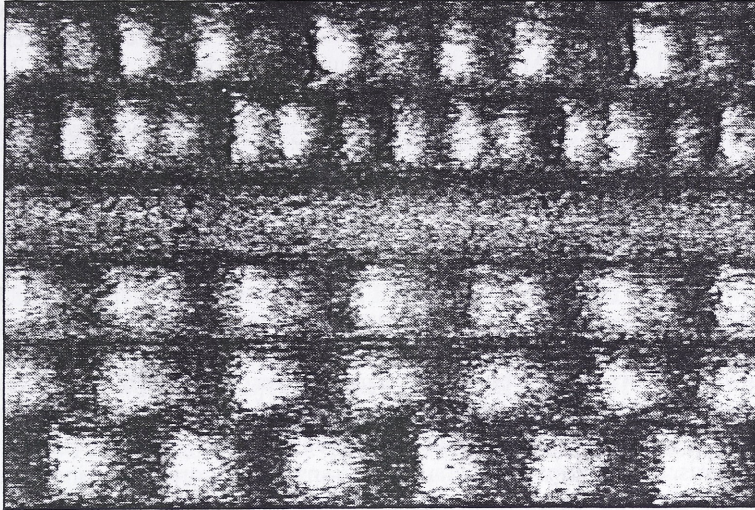


Abb. 3: Draufsicht auf die Tonspuren auf der Innenseite der Galvanos

Zur berührungslosen Vermessung der Höhenprofile gibt es verschiedene Ansätze aus dem Bereich der 3D-Vermessung, die sich durch unterschiedliche Genauigkeiten, Objektgrößen und Anforderungen an die Objektoberfläche auszeichnen. Im Rahmen der Entwicklungsarbeit von SpuBiTo wurden drei verschiedene Ansätze implementiert und getestet, die von ihren Eigenschaften in die engere Wahl kamen. Die drei Ansätze mit ihren Vor- und Nachteilen werden im folgenden vorgestellt.

## Passive Stereoanalyse

Die Verfahren der *Passiven Stereoanalyse* ermitteln die räumliche Struktur von Objekten ebenso wie der Mensch:

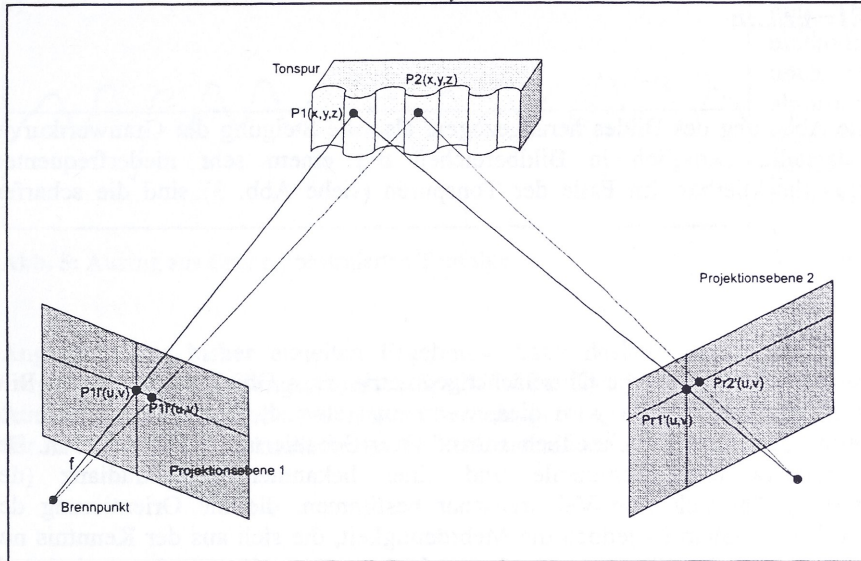


Abb. 4: Prinzip der passiven Stereoanalyse

durch Betrachten einer Szene aus zwei verschiedenen Richtungen. Mathematisch läßt sich dies über den aus der Photogrammetrie bekannten Kollinearitätsansatz beschreiben [SASSE94]. Der Ansatz beruht auf der Annahme, daß ein zu messender 3D-Punkt von 2 oder mehr Kameras, die eine bekannte geometrische Anordnung zueinander aufweisen, wahrgenommen wird. Die Kollinearitätsbedingung besagt, daß Objektpunkt, Bildpunkt und Kamera Brennpunkt für jede Kamera auf einer Gerade liegen müssen. Diese Gerade wird durch 2 Ebenen beschrieben, die parallel zur x- und zur y-Richtung der Bildebene liegen. Wird nur eine Kamera verwendet, so läßt sich nicht bestimmen, an welcher Stelle der Geraden sich der Objektpunkt befindet. Werden zwei Kameras verwendet,

ergibt sich das Gleichungssystem

$$\begin{aligned}
 I) & (a_{11} - x_l \cdot a_{41}) \cdot x_{3D} + (a_{12} - x_l \cdot a_{42}) \cdot y_{3D} + (a_{13} - x_l \cdot a_{43}) \cdot z_{3D} + (a_{14} - x_l \cdot a_{44}) = 0 \\
 II) & (a_{21} - y_l \cdot a_{41}) \cdot x_{3D} + (a_{22} - y_l \cdot a_{42}) \cdot y_{3D} + (a_{23} - y_l \cdot a_{43}) \cdot z_{3D} + (a_{24} - y_l \cdot a_{44}) = 0 \\
 III) & (b_{11} - x_r \cdot b_{41}) \cdot x_{3D} + (b_{12} - x_r \cdot b_{42}) \cdot y_{3D} + (b_{13} - x_r \cdot b_{43}) \cdot z_{3D} + (b_{14} - x_r \cdot b_{44}) = 0 \\
 IV) & (b_{21} - y_r \cdot b_{41}) \cdot x_{3D} + (b_{22} - y_r \cdot b_{42}) \cdot y_{3D} + (b_{23} - y_r \cdot b_{43}) \cdot z_{3D} + (b_{24} - y_r \cdot b_{44}) = 0
 \end{aligned} \tag{1}$$

wobei  $x_{3D}$ ,  $y_{3D}$  und  $z_{3D}$  die Koordinaten des gesuchten Punktes sind und  $x_l$ ,  $y_l$  bzw.  $x_r$ ,  $y_r$  die Bildpunktkoordinaten der beiden Kameras. Der 3D-Punkt wird also durch ein überbestimmtes lineares Gleichungssystem beschrieben, das durch ein geeignetes numerisches Verfahren wie den Householder-Algorithmus fehlerminimierend gelöst werden kann.

Ein Problem dieses Verfahrens ist jedoch die eindeutige Zuordnung der Bildpunkte zueinander. Die 3D-Koordinaten eines Punktes können nur zuverlässig berechnet werden, wenn die verwendeten 2D-Koordinaten der Punkte in den



Bildern auch wirklich denselben Punkt beschreiben. Eine Möglichkeit zur Lösung dieses Problems ist die Verwendung von Lichtquellen, die zur Identifizierung der Objektbereiche ein strukturiertes Lichtmuster auf die Objekte projizieren. Aufgrund der vorliegenden Größenverhältnisse schied dieser Ansatz jedoch aus. Statt dessen werden die korrespondierenden Bildbereiche durch ein in [KLETTE94] beschriebenes Block-Matching-Verfahren gefunden. Das Verfahren stützt sich auf eine akkumulierte Grauwertkorrelation. Als übereinstimmend werden Bildbereiche angesehen, die den höchsten Korrelationswert aufweisen.

## Shape From Defocus

Ein weiterer vielversprechender Ansatz ist die Messung durch gezielte Defocussierung, das sog. *Shape From Defocus*. Hier nutzt man einen Umstand aus, der normalerweise als eher lästig empfunden wird: die geringe Schärfentiefe von Mikroskopobjektiven. Es ist im allgemeinen nicht möglich, ein Objekt mit signifikanter Tiefenausdehnung im Mikroskop komplett scharf zu erkennen, da die Schärfentiefe des Objektivs gegenüber der Objektausdehnung klein ist. Bei bekannten Eigenschaften des Objektivs weisen die scharfgestellten Punkte im Bild einen definierten Abstand zum Aufnahmesystem auf und ergeben damit einen rechnerisch auswertbaren Tiefenwert. Wenn ein Algorithmus eingesetzt wird, der in der Lage ist, scharf gestellte Bildpunkte zu detektieren, kann durch definiertes Verfahren der Aufnahmeeinrichtung in z-Richtung eine Tiefenkarte berechnet werden. Da die Unschärfe in Bildbereichen die gleichen Eigenschaften aufweist wie eine Tiefpaßfilterung, können die scharf gestellten Bereiche durch eine Hochpaßfilterung detektiert werden. Nach [MÜLLER96] ergibt sich ein Tiefenwert  $T(x,y)$  durch

$$T(x,y) = k, \text{ wenn } \frac{|I_x(x,y)|^2}{I_x(x,y)} > \frac{|I_j(x,y)|^2}{I_j(x,y)}, i=1,2,\dots,n \quad (2)$$

Als Hochpaßkriterium wird hier die erste Ableitung des Bildes herangezogen, also die Steigung der Grauwertkurve, was ein hinreichend genaues Maß darstellt. Lediglich in Bildbereichen mit einem sehr niederfrequenten Grauwertverlauf ist die Tiefe nicht sicher detektierbar. Im Falle der Tonspuren (siehe Abb. 3) sind die scharfen Bildbereiche jedoch gut detektierbar.

## Shape From Shading

Der Ansatz des *Shape From Shading* beruht auf der Idee, die Oberflächengeometrie eines Objektes aus der im Bild sichtbaren Schattierung zu rekonstruieren. Nachvollziehbar wird dies, wenn man etwa das Foto einer Gipsbüste betrachtet, bei dem man die räumliche Gestalt der Büste ausschließlich anhand ihrer Schattierung erkennen kann. Bei einem bekannten Winkel zwischen Kamera und Lichtquelle und einer bekannten Objektradianz (der Rückstrahlungsfähigkeit des Objektmaterials) läßt sich eine Vektorenschar bestimmen, die die Orientierung der Oberfläche in einem Bildpunkt beinhaltet. Das Problem ist jedoch die Mehrdeutigkeit, die sich aus der Kenntnis nur einer Beleuchtungsrichtung ergibt. Diese kann durch den Einsatz von mehreren Lichtquellen gelöst werden, wobei auch die Abhängigkeit von der Objektradianz vermieden wird. Die Orientierung eines Oberflächenpunktes ergibt sich nach [KLETTE96] dann durch

$$\vec{n} = \left( \frac{E_{01}}{E_{03}} \cdot E_2 \|s_2\| s_1 - \frac{E_{02}}{E_{03}} \cdot E_1 \|s_1\| s_2 \right) \times \left( \frac{E_{01}}{E_{03}} \cdot E_3 \|s_3\| s_1 - E_1 \|s_1\| s_3 \right) \quad (3)$$

wobei  $E_{01}..E_{03}$  die Grauwerte der Aufnahmen,  $E_1..E_3$  die Beleuchtungsstärken und  $s_1..s_3$  die Beleuchtungsrichtungen der Lichtquellen sind. Es gibt neben diesem Ansatz in der Literatur noch zahlreiche verschiedene Shape-From-Shading-Ansätze, die unterschiedliche Objekteigenschaften berücksichtigen und unterschiedliche Einschränkungen haben, aber der vorgestellte Ansatz wird als der momentan aussichtsreichste zur Rekonstruktion der Spurrhöhen angesehen.



## Ergebnisse und Ausblick

Während der Entwicklungsphase konnten zahlreiche Probleme, die in optisch abtastenden Verfahren auftreten, wie z.B. die automatische Nachführung der Spuren und die Kompensation der vorhandenen Geometriefehler der Galvanos, zufriedenstellen und robust gelöst werden. Dies eröffnete die Chance, die oben beschriebenen Algorithmen zur Tiefenvermessung zu implementieren und auf realen Bildern zu testen. Es wurde dabei Wert auf die Austauschbarkeit der verschiedenen Verfahren nach dem Baukastensystem gelegt, um ein Verfahren einfach durch ein geeigneteres austauschen zu können und so eine hohe Wartbarkeit der Software zu gewährleisten.

Um die Qualität der gewonnenen Klänge vergleichen zu können, wurde ein Galvano gewählt, für den ein abspielbares Positiv zur Verfügung stand, und daraus eine konventionell abgespielte Referenz-Tonfolge gewonnen. Der Vergleich des konventionell Abgespielten mit dem durch SpuBiTo Gewonnenen erbrachte den Nachweis der Äquivalenz, sowohl analytisch wie auch subjektiv durch den Menschen. Das bedeutet, daß das System in der Lage ist, auf optischem Wege Tonfolgen zu gewinnen, die mit den echten übereinstimmen. Abb. 5 zeigt beispielhaft einen Auszug aus einer nach dem Shape-From-Shading-Ansatz gewonnenen Klangkurve. Außerdem wird der Autor im Rahmen des EVA-Vortrages Auszüge aus beiden Tondokumenten zum Vergleich abspielen.

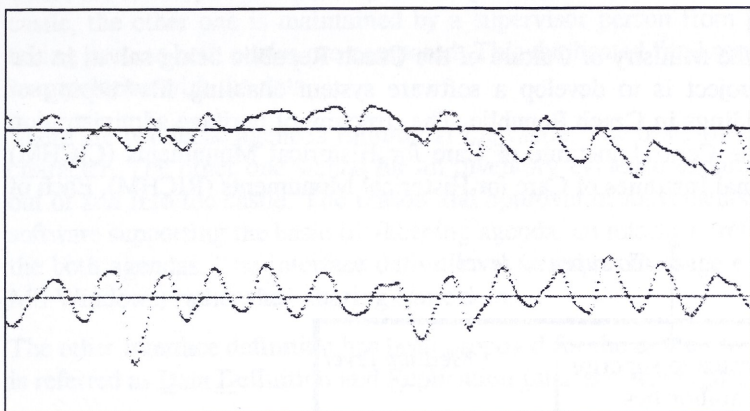


Abb. 5: Auszug aus einer rekonstruierten Tonfolge

Die zum jetzigen Zeitpunkt gewonnenen Klänge erreichen in ihrer Dynamik und ihrem Kurvenverlauf noch nicht die Qualität der konventionell gewonnenen Aufnahmen. Bemerkbar wird dies über einen verwaschenen Klang und ein starkes Hintergrundrauschen. Für eine gute Abtastung, die die Qualität der Originalmethode erreicht oder sogar überschreitet, muß noch mehr Entwicklungsarbeit in die Abtastalgorithmen gesteckt werden. Voraussichtlich wird das endgültige System auf ein Verfahren gestützt sein, das eine Kombination von mehreren der vorgestellten Ansätzen darstellt. Hier steckt noch ein Entwicklungspotential, das 1998 ausgeschöpft werden wird.

Angesichts der bisher erzielten Ergebnisse kann deshalb davon ausgegangen werden, daß am Schluß des hier vorgestellten Entwicklungsprojektes SpuBiTo ein System zur Verfügung steht, mit dem die bisher unzugänglichen Aufnahmen auf den Galvanos in guter bis sehr guter Qualität abgespielt werden können. Die Wieder-Nutzbarmachung der historischen Tondokumente des Völkerkundemuseums ist damit in greifbare Nähe gerückt.

## Literatur

- [EXNER97] Exner, Nicolai: „Phonographenwalzen: Probleme bei der Restaurierung, Informationssicherung und -übertragung von Tonträgern aus Wachs“, Diplomarbeit an der Fachhochschule für Technik und Wirtschaft Berlin, 1997
- [STANKE97] Stanke, Gerd, Wöhrle, Tim: „Bildanalytische Gewinnung von Tonsignalen aus alten Tonträgern - Kupfernegative von Edisonwalzen“, Band „Einsatz neuer Technologien in den Geisteswissenschaften“ S.48-50, Forschungszentrum Jülich 1997
- [ZIEGLER95] Ziegler, Susanne: „Die Walzensammlungen des ehemaligen Berliner Phonogramm-Archives“ In: Baessler-Archiv, Beiträge zur Völkerkunde, Band XLIII, Sonderdruck aus Heft 1, Berlin 1995
- [SASSE94] Sasse, Ralph: „Bestimmung von Entfernungsbildern durch aktive stereoskopische Verfahren“, Vieweg Verlag Braunschweig, 1994
- [MÜLLER96] Müller, R.: „Dreidimensionale Bildrekonstruktion aus Serien lichtmikroskopischer Aufnahmen“, International Journal for Light and Electron Optics, Stuttgart 1996
- [KLETTE96] Klette, Reinhard et.al.: „Computer Vision-räumliche Information aus digitalen Bildern“, Vieweg Verlag, Braunschweig 1996