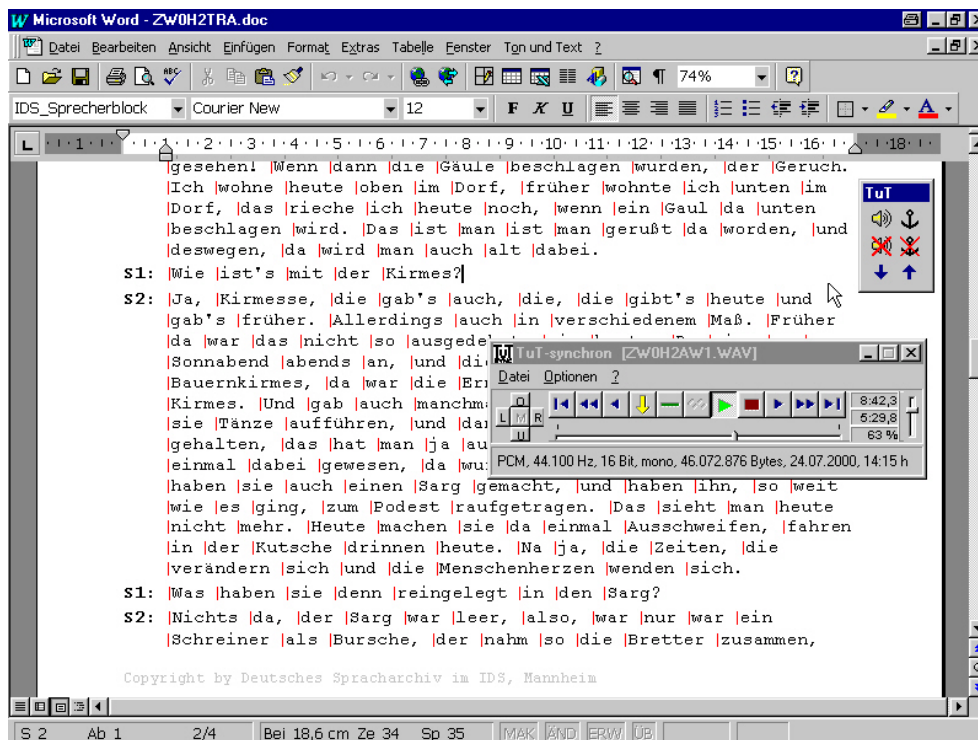


## "Ton und Text" - ein Tonwiedergabe-Programm für Transkribenten

Wolfgang Schneider

**Kurzbeschreibung:** Beim Transkribieren findet - teils absichtlich, teils technisch bedingt - ein Medienwechsel vom Ton zur Schrift statt. Damit daraus kein unüberwindlicher Medienbruch wird, sind Hilfsmittel für das Transkribieren erforderlich, die die Integration der beiden Medien unterstützen. Mit den beiden Programm-Versionen von "Ton und Text" (für Windows-PCs) lassen sich Tonaufnahmen und deren Verschriftlichungen effizient neben- und miteinander nutzen: Beim Erstellen des Transkripts kann man mit den TuT-Programmen neben dem Schreiben den Ton gezielt wiedergeben, und unter bestimmten Voraussetzungen ist mit TuT-synchron sogar die text-synchrone Wiedergabe der Tonaufnahme aus dem Transkript heraus möglich.

**Bezugsquelle:** Die aktuellen Versionen der TuT-Programme und weitere Hilfsmittel für Transkribenten stehen im Internet unter [www.Ton-und-Text.de](http://www.Ton-und-Text.de) zum kostenlosen Download bereit. Da die Programme ständig weiterentwickelt werden (siehe den letzten Abschnitt des Beitrags), lohnt sich die regelmäßige Überprüfung auf Neuigkeiten.



## Medienwechsel beim Transkribieren

Das Transkribieren gesprochener Sprache ist ein Verfahren, das flüchtige, gesprochene Wort unter Zuhilfenahme von Tonaufzeichnungen nachträglich in schriftlicher Form zu fixieren. Das Resultat, ein Transkript, ist in der Regel nicht - wie beim Diktat - die schriftliche Fassung eines gesprochenen Textes, sondern mehr: Beim Transkribieren wird das Gehörte analysiert bzw. für eine Analyse aufbereitet. Zusätzlich zur Textfassung des Gesprochenen werden die für die Analyse wichtigen Momente festgehalten. Nicht selten geht es dabei von vornherein um eine spezifische wissenschaftliche Fragestellung, die die Aufmerksamkeit des Transkribenten auf ganz bestimmte Phänomene lenkt. Egal, wie spezifisch dieses Interesse ist, zum Transkribieren gehört immer die Anwendung elaborierter, konventionalisierter Schreibweisen, schon um die (Un-)Gleichzeitigkeit des Sprechens zu verdeutlichen. Wer Intonation dokumentieren will, muß zu Annotationen greifen, die eingeführt sind. Wer Textstellen im verschrifteten Diskurs kommentieren will, kann ebenfalls auf eingeführte technische Verfahren zurückgreifen. Ein Reihe von ausgearbeiteten und systematisierten Verfahrensvorschlägen liegt vor, z.B. HIAT<sup>1</sup>, DIDA des IDS<sup>2</sup>, GAT<sup>3</sup> etc. Einen Überblick über diverse Transkriptionsverfahren bietet Redder.<sup>4</sup>

Gleichgültig, nach welchen Konventionen gesprochene Sprache transkribiert wird, in jedem Falle erfolgt dabei absichtlich ein *Wechsel des Mediums* vom Ton zum Text. Das ist schon deshalb unabdingbar, da sich Annotationen, Kommentare und Analysen nur in der Schriftform ohne Beeinträchtigung des Gesprochenen diesem zuordnen lassen. Annotationen nachträglich zur originalen Tonaufzeichnung z.B. auf eine zweite (parallele) Tonspur aufzusprechen - wie beim Verfahren des "Kino für Blinde" -, würde bei der Wiedergabe zwangsläufig als Störung der originalen Tonaufzeichnung wahrgenommen - einmal abgesehen von diversen praktischen Schwierigkeiten der Synchronisation, da der Umfang der Annotationen in keinem proportionalen Verhältnis zur Dauer der zu annotierenden Ereignisse steht.

Die Tätigkeit des Transkribenten besteht also darin, gesprochene Sprache in ein anderes Medium zu übertragen. Transkriptionskonventionen bieten Anleitungen dafür, wie die Höreindrücke in der Gestaltung des Transkripttextes zu fixieren sind, damit der Leser (auch der Transkribent selber, in seiner Funktion als Analy-

<sup>1</sup> HalbInterpretative ArbeitsTranskriptionen (HIAT) ist der Name eines Transkriptionsverfahrens, das von K. Ehlich und J. Rehbein (siehe: Linguistische Berichte, Nr. 45, 1976 und Nr. 59, 1979) entwickelt wurde. Die praktische Realisation dieser Konventionen wird beim Transkribieren am Computer mit den Transkriptionseditoren HIAT-DOS (ursprünglich für DOS-Rechner entwickelt, ist es aber immer noch auch mit Windows-Rechnern problemlos verwendbar) und mit SyncWRITER (für MACs) unterstützt. Einer kurze Darstellung des HIAT-Verfahrens und Hinweise zu Werkzeugen für HIAT-Transkribenten befinden sich unter: <http://www.DaF.Uni-Muenchen.de/HIAT>.

<sup>2</sup> Informationen zum Diskursdatenbank-System DIDA und zu den Transkriptionskonventionen des IDS sind unter <http://www.IDS-Mannheim.de/prag/dida> nachzulesen.

<sup>3</sup> Siehe: Selting / Auer / Barden u.a. (1998): Gesprächsanalytisches Transkriptionssystem (GAT). In: Linguistische Berichte 173, 91-122; siehe auch: <http://www.fb1s.uni-hannover.de/sdls/schlobi/schrift/GAT/>.

<sup>4</sup> Siehe A. Redder (2001): Aufbau und Gestaltung von Transkriptionssystemen. In: Handbücher zur Sprach- und Kommunikationswissenschaft, Band 16.2, "Text und Gesprächslinguistik", Kapitel "XV. Methoden II: Transkription". Berlin: de Gruyter.

sierender) durch die Gestaltung des Schriftmediums später bei der gedanklichen Rekonstruktion des originalen Sprechereignisses unterstützt wird. Häufig soll z.B. im Transkript das Gesprochene parallel zum Zeitablauf des Ereignisses in der Schriftform "begleitet" werden, um für den Leser den zeitlichen Ablauf deutlich werden zu lassen (natürlich nur im Rahmen der für den jeweiligen Analysezweck erforderlichen Genauigkeit). Die Textform wird deshalb oft entlang einer imaginierten Zeitachse gestaltet, um nicht nur das zeitliche Nacheinander, sondern vor allem auch das gleichzeitig Gesprochene deutlich erkennbar werden zu lassen. Hier gibt es die unvermeidliche Schwierigkeit, dass der Platzbedarf für die Schriftform (auch ohne die Annotationen) kein enges proportionales Verhältnis zur Zeitdauer des Gesprochenen hat, dass das Transkript also nicht entlang einer Zeitskala mit äquidistanter Metrik notiert werden kann, wenn man das schriftliche Abbild nicht bis zu Unleserlichkeit verzerren will.

Durch den Wechsel des Mediums vom Ton zum Text - und nur dadurch - kann zudem die scheinbar paradoxe Leistung gelingen, das nacheinander Gesprochene quasi gleichzeitig be(tr)achten zu können (wenn auch nicht gleichzeitig zu hören), da geübte Leser einmal gelesenen Text bei einer wiederholten Lesung auch überblicksartig rezipieren können. Im Transkript kann man sich also zeitlich relativ weit auseinanderliegende Sprechereignisse gleichzeitig vor Augen führen.

Mit dem Übergang von der Tonaufzeichnung zum Transkript wird nicht nur ein absichtlicher Medienwechsel vollzogen. Dieser Wechsel gerät nun unvermeidlich zu einem Medienbruch, zum Zerlegen des festgehaltenen Gesprochenen in Tonaufnahme und Transkript, die im Fortgang der Analyse und der Präsentation der Ergebnisse nur recht umständlich wieder aufeinander bezogen werden können.

## **Beide Medien am Computer**

Von der Tonaufzeichnung, die auf Magnetbändern konserviert und tradiert wurde, erstellte der Transkribent früher die hand- oder maschinenschriftliche Aufzeichnung. Die Tonwiedergabe erfolgte vom Band, das Abschnitt für Abschnitt in mehreren Wiederholungen abgespielt wird, und die schriftliche Erfassung des Gehörten erfolgte - technisch davon völlig getrennt - meist in *einem* sehr langsamen Durchgang. Mit der Verwendung einer computer-gestützten Textverarbeitung ändert sich die Arbeit des Transkribenten in einem wesentlichen Punkt. Die Textverarbeitung effektiviert die Transkriptionsarbeit insofern, als nun die Transkription in *mehreren* Durchgängen erfolgen kann. Korrekturen und Annotate lassen sich nachträglich an jeder Stelle ohne Doppelarbeit einarbeiten. Die Konzeption der Arbeitstranskription, die für das HIAT-Verfahren namensstiftend ist, also der prinzipiell auf wiederholte Arbeitsschritte angelegte Transkriptionsprozess, hat im Computer ihr kongeniales Werkzeug. Da beim Transkribieren mit Standard-Textverarbeitungsprogrammen prinzipielle Strukturprobleme auftreten - diese Editoren unterstützen das Schreiben in mehreren parallelen Spuren (Partiturschreibweise) nicht -, fällt die Effizienzsteigerung mit Word & Co. jedoch geringer aus, als sie theoretisch sein könnte.

Mit der Digitalisierung von Tonaufnahmen wird die Voraussetzung geschaffen, auch die Tonaufzeichnung mit Mitteln der Datenverarbeitung zu erschließen und zu behandeln. Durch den Wechsel des Datenträgers für die Tonaufzeichnung vom

Magnetband (gleichgültig ob mit traditioneller analoger oder modernerer digitaler DAT-Aufzeichnung) zu einem Direktzugriffsspeicher wird zudem erstmals eine völlig neue Art des Zugriffs auf die Tonkonserve möglich. Einzelne Stellen des originalen Ereignisses müssen vor der Wiedergabe nicht mehr zeitraubend durch Vor- oder Rückspulen auf einem sequentiellen Magnetband angesteuert werden, sondern lassen sich (fast) verzögerungsfrei von Mini-Disk oder CD-Audio oder - noch weitergehend in ein Computersystem integriert - von einem Dateisystem auf CD-ROM oder Festplatte "direkt" auslesen und wiedergeben.

### **Konzepte zur Verknüpfung beider Medien**

Die beiden Speicherformen gesprochener Sprache, das Transkript und die Tonaufzeichnung, existieren nun zunächst, auch wenn sie auf demselben Computer verwaltet werden, als unverbundene Medien nebeneinander. Das bedeutet, dass die zum jeweils Gehörten gehörige Stelle im Transkript nicht unmittelbar zu sehen ist, und das im Transkript Gelesene in der Tonaufnahme nicht ohne weiteres auffindbar ist. Die beiden unverbundenen Medien dauerhaft eng aneinander zu binden ist die Funktion spezialisierter Programme. Sie lassen sich kategorisieren nach zwei prinzipiell unterschiedlichen Vorgehensweisen:

- Das erste Verfahren kommt vorwiegend bei phonetischen oder phonetisch-phonologischen Analysen gesprochener Sprache, die sich sehr detailliert auf kurze Ereignisse konzentrieren, zum Einsatz: Für diesen Bereich existieren Programme (z.B. praat<sup>5</sup>), die es erlauben, digitalisierte Tonaufnahmen - ähnlich wie in Ton-Editoren - zu visualisieren (Sonagramme, Frequenzspektren, Grundfrequenz etc.), jede Stelle wiederzugeben und zu annotieren. In "label-tiers" lassen sich alle Bestandteile des Transkripttextes - aber auch Annotate - wie Etiketten quasi anheften an die Visualisierung des Gesprochenen. Die graphische Repräsentation der Tonaufnahme gibt dabei das Gerüst vor, in dem der transkribierte Text zeitlich recht exakt (wortweise oder gar silbenweise genau) angeordnet werden kann. Diese Art der originalgetreuen Transkription ist jedoch nur für kurze Ausschnitte sinnvoll verwendbar, da der Text dabei nicht als kontinuierliche Wortfolge notiert ist und deshalb das flüssige oder überblicksartige Lesen des Textes recht mühsam wird.
- Das andere Verfahren knüpft an die üblichen Transkriptionsverfahren an, die für die Wiedergabe diskursiver Einheiten besser geeignet sind, schon bei Tonaufnahmen, die nur wenige Minuten umfassen. Es initiiert deshalb die Verbindung mit dem Medium Ton vom Text aus. Nicht die Tonaufzeichnung bildet hier das Gerüst für die Verknüpfungen zum Transkript, sondern vom Transkript aus wird auf die zugehörigen Stellen in der Tonaufzeichnung verwiesen. Das Programm "Ton und Text"-synchron realisiert diese Art der Verknüpfung der beiden Medien.

---

<sup>5</sup> Siehe: <http://www.praat.org>.

## Zugriff auf die Tonaufnahme beim Transkribieren

"Ton und Text" wurde ursprünglich 1996 entwickelt, um eine stabile und effiziente Tonwiedergabe digitalisierter Aufnahmen aus den Korpora am Lehrstuhl von Gisela Brüner (Universität Dortmund) auf handelsüblichen PCs zu ermöglichen. Zum damaligen Zeitpunkt versagten die für Standard-PCs verfügbaren Werkzeuge schon bei Aufnahmen von wenigen Sekunden den Dienst. Sogar die teuren Ton-Editoren waren schon bei Aufnahmen von 10 Minuten und länger überfordert bzw. arbeiteten dann nur noch recht zähflüssig. Unter Verzicht auf die Leistungen ausgewachsener Tonstudio-Programme und durch Konzentration auf die effiziente Tonwiedergabe gelang es, ein bequem zu handhabendes Abspielprogramm zu entwickeln, das auch mit ungewöhnlich langen Aufnahmen ohne gesteigerten Ressourcenbedarf umgehen kann.

Zwar ist das alles inzwischen Geschichte, die Standard-Computerausstattung (Software und Hardware) ist leistungsfähiger geworden und erlaubt heute den Umgang mit stundenlangen Aufnahmen. Die Auswahl an *Playern* für Tonwiedergabe ist so umfangreich wie noch nie. Dennoch ist die Weiterentwicklung eines Wiedergabeprogramms speziell zum Transkribieren auch heute noch sinnvoll, um bestimmte Leistungen zu ermöglichen, die alle anderen bisher bekannten Abspielprogramme vermissen lassen. Es existieren bisher zwei Versionen von "Ton und Text", die sich nur im Funktionsumfang unterscheiden. Das einfachste von ihnen, das Abspielprogramm TuT-solo bietet:

- *Effiziente Navigation im Ton durch Sprünge variabler Länge*: Es werden kurze und weite Sprünge unterschieden, deren Sprungweite vom Anwender nach Bedarf vorgegeben werden kann. Damit läßt sich jeder Zeitpunkt innerhalb einer Aufnahme effizient und präzise ansteuern.
- *Wiederauffinden von Tonstellen durch "Zeitpunkte-Memory"*: An interessanten Zeitpunkt-Positionen können "Memory"-Anker gesetzt werden, zu denen man später sehr schnell zurückfinden kann. Die Positionen werden weder im Transkript, noch in der Tonaufnahme vermerkt. Sie bleiben automatisch für die 6 zuletzt verwendeten Aufnahmen im System erhalten.
- *Endlos-Wiederholungen von frei wählbaren Tonausschnitten mit definierbarer Pause zwischen den Wiederholungen*: Jeder Ausschnitt, der wiederholt wiedergegeben werden soll, kann sehr präzise definiert (dazu ist während der Dauer der Wiedergabe des Ausschnitts eine Taste gedrückt zu halten) und anschließend beliebig oft wiederholt werden. Um den Höreindruck am Anfang und Ende des Ausschnitts nicht zu stören, läßt sich eine frei definierbare Pause zwischen den Wiederholungen einlegen. Die Wiederholung kann bei Bedarf aber auch ohne Pause starten. Unverständliche Passagen lassen sich auf diese Weise viel besser aufklären, als das bei wiederholter Wiedergabe mit traditionellen Tonbandgeräten, die zwischen den Wiederholungen zeitraubend zurückspulen müssen, möglich ist. So trägt dieses Hilfsmittel dazu bei, die Qualität der Transkription zu verbessern.
- *Steuerbarkeit aller Funktionen wie üblich über Icons oder Menü per Maus aber auch per Tastatur-Kürzel*: Für die persönlichen Arbeitsgewohnheiten stehen unterschiedliche Steuerungsmöglichkeiten zur Verfügung. Das steigert die Effizienz der Anwender.

TuT-solo läßt sich beim Transkribieren mit jedem Editor verwenden. Es effektiviert nur den Zugriff auf den Ton, funktioniert aber unverbunden, im Prozess der Erstellung des Transkripts. Es läßt sich also auch neben HIAT-DOS, das unter Windows in einer DOS-Box läuft, einsetzen.

## **Synchronisierung von Transkript und Tonaufnahme**

Mit dem Programm TuT-synchron läßt sich der Ton mit dem Transkripttext gezielt verknüpfen. Es beinhaltet alle Leistungsmerkmale von "Ton und Text"-solo und bietet zusätzlich:

- *Fernsteuerbarkeit aller Funktionen aus anderen Programmen*, insbesondere Textverarbeitungen. Vorbereitet wurde die Steuerbarkeit des Wiedergabeprogramms aus (fast) allen makro-fähigen Textverarbeitungen unter Windows (Für Techniker: Die Fähigkeit zur DDE-Kommunikation aus Makros ist dazu erforderlich). Auch ein spezialisierter Transkriptionseditor - so es ihn denn unter Windows einmal gibt - könnte daran anknüpfen. Die "Marktlage" hat jedoch dazu geführt, dass bisher tatsächlich nur die Ansteuerung aus Microsoft Word heraus realisiert wurde (die Makros der aktuellen TuT-synchron-Version funktionieren ab Word 97).

Um die Fernsteuerbarkeit von TuT-synchron nutzen zu können, müssen die Transkripte im Format von Word vorliegen. Das dürfte zwar nicht immer von Anfang an der Fall sein, aber in der Regel lassen sich "fertige" Transkripte nachträglich in das Word-Format konvertieren. Das Word-Transkript wird mit einer besonderen Druckformatvorlage verbunden, die die TuT-Makros enthält, mit denen vom Word-Transkript aus die Verbindung zum Tonwiedergabeprogramm TuT-synchron erfolgt. An den interessanten Stellen (Sprecherwechsel, Satzanfänge etc.) werden manuell in das Transkript Synchronisationsanker gesetzt. Synchronisationsanker bestehen aus einer unsichtbaren Zeitangabe und einer schmalen, nur bei Bedarf sichtbaren Markierung. Das Setzen der Anker erfolgt an der aktuellen Position des Textcursors. Während der Tonwiedergabe ist zu genau dem Zeitpunkt, an dem der Ton der verschrifteten Stelle zu hören ist, eine Taste zu drücken. Um die unvermeidlichen kleinen Fehler, die bei der manuellen Arbeitsweise entstehen (Reaktionszeit), zu korrigieren, kann man die Positionierung des Ankers in einer Visualisierung des Tons mit der Maus präzise nachbessern. Die Dichte und Zahl der Synchronisationsanker ist im Prinzip beliebig. Sie ist in Abhängigkeit vom Untersuchungszweck und von der verfügbaren Bearbeitungszeit zu wählen. Sofern die Zeitpunkte nicht manuell gesetzt werden müssen, sondern durch ein automatisches Alignment aus der Wortliste des Transkripts und der Tonaufzeichnung errechnet werden, wie es am IDS praktiziert wird,<sup>6</sup> kann die Synchronisierung sogar wortweise erfolgen. TuT-synchron stellt nur die Möglichkeit zur Verfügung, die Synchronanker manuell zu setzen. Automatisch erzeugte Synchronanker können jedoch verwendet werden.

Wenn ein Transkript mit Synchronisationsankern versehen wurde, läßt sich der Ton, der zu der Stelle gehört, an der der Textcursor steht, wiedergeben. Dazu wird mit einem Tastendruck der nächste vorausgehende Synchronisationsanker ge-

---

<sup>6</sup> Siehe: <http://www.IDS-Mannheim.de/prag/alignment.html>.

sucht, der unsichtbar vermerkte Zeitpunkt ausgelesen und TuT-synchron (per DDE) beauftragt, einen Tonausschnitt wiederzugeben. Das kann - je nach vorgewählter Option - ein Intervall fester oder variabler Länge sein, es können aber auch noch Vorlauf- und Nachlaufintervalle dabei sein, damit auch der Ton in der näheren Umgebung der angeforderten Stelle zu hören ist.

Mit TuT-synchron lassen sich das Word-Transkript und die digitalisierte Tonaufnahme als Medien-Verbund nutzen. Der schnellen Navigation im Transkript, die mit dem Blättern im Text oder der Suchfunktion möglich ist, folgt mittels der Synchronanker die ebenso schnelle und präzise Navigation in der Tonaufnahme. Insbesondere beim Hörvergleich ähnlicher Textstellen ist dies von großem Vorteil. Man kann sehr schnell von einer Fundstelle zur nächsten und zurück springen und ebenso schnell den zugehörigen Ton hören. Es ist erstaunlich, wie sehr das Vergleichen distanter Gesprächsstellen erleichtert wird, wenn der Höreindruck nicht durch die Wartezeit und durch das auditive Navigieren irritiert wird.

### **Was noch zu tun ist**

Für die Weiterentwicklung der TuT-Programme ist (neben der Änderung technischer Interna, die die Kompatibilität mit zukünftigen Windows-Versionen garantieren sollen) vorgesehen die Unterstützung

- beim *Schneiden von Ausschnitten* aus der Tonaufnahme.
- beim *Übergang zu anderen Werkzeugen* (z.B. praat) zur Visualisierung und Analyse von Tonausschnitten, z.B. für phonetische Untersuchungen.
- bei *Tonüberblendungen zur Anonymisierung*. Dabei wird ein Verfahren eingesetzt, das schneller ist, weil es mit wesentlich bescheidenerer Rechner-Leistung auskommt, als dies bei echten Toneditoren üblich ist.
- bei der Anwendung *spezieller Transkriptionskonventionen*. Transkripte in Partitur-Schreibweise erfordern (im Word-Format) besondere Rücksichtnahmen bei der Darstellung der Synchronanker - die müssen sich über alle Zeilen einer Partitur erstrecken.
- bei der *synchronen Verknüpfung von Filmaufnahmen* mit Transkripten. Das war konzeptionell immer vorgesehen und ist von mir in Experimenten auch schon erfolgreich praktiziert worden.

Aus den Word-Transkripten werden dann Integrationszentren für alle Medien, die neben der Tonaufnahme zur Rekonstruktion des originalen Sprechereignisses und zur Analyse hilfreich sind, wie Film oder auch statische Bebilderungen (Fotos, Skizzen oder ausschnittweise Visualisierungen der Tonaufzeichnung).

Wolfgang Schneider  
info@Ton-und-Text.de

Veröffentlicht am 10.9.2001

© Copyright by GESPRÄCHSFORSCHUNG. Alle Rechte vorbehalten.