

## Das Stilisieren von Timing-Konturen

**Citation for published version (APA):**

Adriaens-Porzig, U., & Rump, H. H. (1990). *Das Stilisieren von Timing-Konturen: eine neue Methode zur Untersuchung temporaler Strukturen in gesprochener Sprache*. (IPO rapport; Vol. 762). Instituut voor Perceptie Onderzoek (IPO).

**Document status and date:**

Gepubliceerd: 14/09/1990

**Document Version:**

Uitgevers PDF, ook bekend als Version of Record

**Please check the document version of this publication:**

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

**General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.tue.nl/taverne](http://www.tue.nl/taverne)

**Take down policy**

If you believe that this document breaches copyright please contact us at:

[openaccess@tue.nl](mailto:openaccess@tue.nl)

providing details and we will investigate your claim.

Rapport no. 762

Das Stilisieren von  
Timing-Konturen

U. Adriaens-Porzig  
H.H. Rump

# Das Stilisieren von Timing-Konturen

Ursula Adriaens-Porzig      Willem Rump\*

August 1990

Eine neue Methode zur Untersuchung  
temporaler Strukturen in gesprochener Sprache

---

\*Student der Phonetik an der Universität Utrecht, studentische Hilfskraft am IPO

# Contents

<b>1</b>	<b>Einleitung</b>	<b>5</b>
1.1	Zielsetzung . . . . .	5
1.2	Einführung . . . . .	5
1.3	Bisherige Untersuchungen und Hintergründe . .	5
<b>2</b>	<b>Pilotstudie</b>	<b>7</b>
2.1	Von Timing-Treppen zur Timing-Kontur . . . .	7
2.2	Ein Beispiel . . . . .	8
2.3	Pretest: Manuell erstellte Timing-Konturen . .	9
2.3.1	Das Testmaterial . . . . .	11
2.3.2	Durchführung . . . . .	11
2.3.3	Ergebnisse . . . . .	11
2.3.4	Diskussion und Schlußfolgerungen . . .	14
<b>3</b>	<b>Die Methode</b>	<b>16</b>
3.1	Timing-Stilisierung . . . . .	16
3.1.1	Das Sprachmaterial . . . . .	17
3.1.2	Phonemdauern . . . . .	17
3.1.3	Das Übertragen der Timingstruktur . .	19
3.1.4	Die Stilisierungen . . . . .	19
3.2	Perzeptionstest: Timing-Stilisierungen . . . .	22
3.2.1	Das Testmaterial . . . . .	22
3.2.2	Durchführung des Testes . . . . .	22
3.2.3	Ergebnisse . . . . .	24
3.2.4	Diskussion und Schlußfolgerung . . . . .	27
<b>4</b>	<b>Diskussion</b>	<b>27</b>
<b>5</b>	<b>Literaturverzeichnis</b>	<b>30</b>
<b>6</b>	<b>Anhang A: Die Satzdauerwerte der 4 Versionen im Pretest</b>	<b>31</b>
<b>7</b>	<b>Anhang B: Die Klatt-Tabelle</b>	<b>32</b>



<b>8</b>	<b>Anhang C: Die verwendeten Sotschecksätze im Perzeptionstest</b>	<b>33</b>
<b>9</b>	<b>Anhang D: Die Satzdauerwerte der Versionen im Perzeptionstest</b>	<b>34</b>
<b>10</b>	<b>Anhang E: Beispiele der Stilisierungen und Ableiter</b>	<b>35</b>
<b>11</b>	<b>Anhang F: Testantwortbogen</b>	<b>39</b>

## Abstract

The major part of the research on temporal structures has focussed on the segmental level. Numerous articles report on temporal regularities that have been found, depending on such variables as intrinsic vowel length, consonant environment, degree of opening of the vocal tract, syllabic structure etc.

A new and very different method was found to investigate timing variations in an utterance; it will be briefly described here. With the new method one no longer looks at the segmental level of an utterance but at curves that span the whole utterance. Analogously to the IPO-method that is used in intonation research one can draw stylized timing-contours, that govern the internal timing of an utterance.

## Zusammenfassung

Der Großteil wissenschaftlicher Untersuchungen auf dem Gebiet temporaler Strukturen bezog sich bislang auf die segmentelle Ebene. Zahlreiche Artikel berichten von temporalen Regularitäten, die in Abhängigkeit von Variablen wie intrinsische Vokallänge, Konsonantenumgebung, Grad der Öffnung des Artikulationsorgans, silbische Struktur usw. auftreten.

Hier wird nun eine neue, abweichende Methode vorgestellt, mit der Timing-Variationen in einer gesprochenen Äußerung untersucht werden können. Mit dieser neuen Methode wird nicht länger nur die segmentelle Ebene betrachtet, sondern der Verlauf einer "Timing-Kontur" über die gesamte sprachliche Äußerung verfolgt. Analog zur IPO-Methode der Intonationsforschung, wird eine stilisierte Timing-Kontur gezeichnet, die das interne Timing der Äußerung steuert.

## Samenvatting

In het onderzoek naar temporele verschijnselen in spraak is de aandacht vooral gericht geweest op segmentele duureigenschappen. Er zijn dan ook talrijke artikelen verschenen, waarin gelet werd op intrinsieke vokaalduur, konsonantomgeving, openingsgraad van de mond, syllabestructuur etc.

In dit artikel wordt een nieuwe methode voorgesteld, waarmee duurvariëaties in spraak onderzocht kunnen worden. Deze methode houdt in, dat zinnen niet op het segmentele, maar op het suprasegmentele nivo worden bekeken. Analooq aan de IPO-methode voor het intonatieonderzoek wordt er een zogenaamde 'timing-kontour' ontworpen, die de interne duurverdeling in een uiting regelt.

## Vorwort von Willem Rump

Im Rahmen meiner Tätigkeit als Praktikant am Instituut voor Perceptieonderzoek (IPO) in Eindhoven habe ich mich von Oktober 1989 bis Juli 1990 mit einer neuen Methode zur Untersuchung temporaler Strukturen in gesprochener deutscher Sprache beschäftigt. Innerhalb dieser Arbeit habe ich ein Perzeptionsexperiment durchgeführt.

Ich möchte Herrn Prof. Dr. W. Hess in Bonn für die gebotene Gelegenheit danken, den Test an seinem Institut (IKP) durchzuführen. Auch Herrn Dr. W. Sendlmeier, IKP, Bonn, möchte ich herzlich danken für seine Unterstützung. Mein Dank geht auch an Drs. Wim Peeters, OTS, Universität Utrecht, Mentor meines Praktikums. Frau Drs. Ursula Adriaens-Porzig hat meine Arbeit am IPO begleitet. Ihr großer Enthusiasmus und ihre große Sachverständigkeit haben dazu beigetragen, daß diese Arbeit zustande kam.

# **1 Einleitung**

## **1.1 Zielsetzung**

Hier wird eine neue Methode vorgestellt, bei der wir unabhängig von der segmentellen Information eine Beschreibung des satzinternen Timings erhalten. Diese Beschreibung kann graphisch als "Timing-Kontur" dargestellt werden. Eine Timing-Kontur beschreibt Veränderungen im Satztiming kontinuierlich, sie ist graphisch eine zusammenhängende Kurve.

## **1.2 Einführung**

Der Aufbau dieses Berichtes entspricht in etwa der Arbeitsabfolge, die uns zur Methode des "Timing-Stilisierens" führte. Das erste Kapitel behandelt bisherige Methoden und deren Ergebnisse und führt in die Thematik des Satztimings ein. Das zweite Kapitel beschreibt eine Pilotstudie, nämlich die ersten Ansätze zur neuen Methode. Diese Pilotstudie wird von Ursula Adriaens-Porzig beschrieben, da sie diese Studie durchgeführt hat. In Kapitel drei beschreibt Willem Rump die neue Methode, die Vorgehensweise, und die perzeptive Überprüfung der neuen Methode in einem Experiment. Im vierten Kapitel wird die neue Methode diskutiert.

## **1.3 Bisherige Untersuchungen und Hintergründe**

Die Zeit spielt beim lautsprachlichen Kommunikationsprozeß auf verschiedenen Ebenen eine große Rolle. Bei der Wahrnehmung einer lautsprachlichen Äußerung kann der Hörer unterscheiden, ob die Äußerung schnell, langsam oder normal gesprochen wurde. Wir nennen das Tempowahrnehmung. Bestimmte Teile eines Satzes werden schneller oder langsamer als andere Teile gesprochen, zum Beispiel unter Einfluß von Akzentuierung. Diese mehr oder weniger regelmäßig auftretenden Tempo-Veränderungen werden Satzrhythmus genannt.

Jede sprachliche Äußerung setzt sich aus Lautsegmenten zusammen. Die Summe der Segmentdauerwerte bestimmt die Gesamtdauer der lautsprachlichen Äußerung, so daß also jeder Segmentdauerwert zur Gesamtdauer beiträgt. Wir wissen, daß jeder Sprachlaut nach Dauer variieren kann. Die einzelnen Lautdauerwerte können durch unmittelbare Beobachtung, d.h. impressionistisch, nicht absolut bestimmt werden.

Der Literatur kann entnommen werden, daß sich bereits zahlreiche Untersuchungen mit der Frage nach den Dauerverhältnissen in gesprochener Sprache auseinandergesetzt haben, sowohl in Hinsicht auf die Dauer von



Vokalen als auch auf die von Konsonanten (z.B. Peterson und Lehiste, 1960; Nooteboom, 1972 und 1973; Fischer-Jørgensen, 1964; Elsendoorn, 1984; Klatt, 1987; und andere). Allerdings beziehen sich die meisten Untersuchungen auf die Stellung und Dauer der Vokale und Konsonanten in isolierten Wörtern.

Im folgenden werden wir die wichtigsten Faktoren nennen, die in der Literatur zum Thema Sprachlautdauer vorgefunden werden. Alle diese Ergebnisse beruhen auf einer analytischen Faktorenbestimmung:

- In fließend gesprochener Sprache sind alle Segmentdauerwerte kürzer, als in isoliert gesprochenen Wörtern.
- Die Dauerwerte von Sprachlautsegmenten kann wichtige syntaktische oder semantische Grenzen markieren, unwichtige Segmente werden stark verkürzt oder sogar unterdrückt.
- Die letzte Silbe vor dem Ende einer prosodisch markierten Wortgruppe ist länger, als in anderen Positionen (präpausale oder präfinale Verlängerung).
- Vokale in Silben am Satzende sind länger als in anderen Positionen.
- Akzentuierte Vokale in mehrsilbigen Wörtern sind kürzer als in Monosilben, nichtakzentuierte Vokale sind kürzer als akzentuierte und der Dauerwert eines akzentuierten Vokals in Wortgruppen hängt von der Anzahl der vorangehenden und der folgenden Akzente ab. (Die letzte Beobachtung konnte allerdings bislang nur für Schwedisch bestätigt werden).
- Die Gesamtdauer eines Wortes hängt ab von der Position des Wortes in der Äußerung und vom Satzakzent, das der Sprecher dem Wort beimißt. Unwichtige Wörter werden verkürzt. Zudem sind Vokale in Silben am Ende eines Wortes länger als in anderen Positionen.
- Der Dauerwert eines Vokals hängt von den ihn umgebenden Konsonanten ab, besonders von dem ihm folgenden. So verlängert sich im Englischen und im Niederländischen die Vokaldauer, wenn dem Vokal ein stimmhafter Plosiv folgt. - Eine weitere Beobachtung besagt, daß ein Vokal nicht mehr verkürzt werden kann, wenn sein Minimaldauerwert erreicht ist. Doch kann dann die Verkürzung im darauffolgenden Konsonanten erfolgen.

Es gibt also Faktoren, die auf die Dauer der Sprachlautsegmente verlängernd wirken und es gibt Faktoren, die verkürzenden Einfluß haben. Offensichtlich ist das Problem des Timings vielschichtig und komplex. Es wäre daher wünschenswert, eine einfache, überschaubare Methode zu haben, mit der die Zeitstruktur einer Äußerung untersucht werden kann. Die Stilisierungsmethode richtet sich auf die Frage, ob es möglich ist, Segmentlängen als von Tempo und Rhythmus abhängig und über größere Satzeinheiten (wie zB. Silbe, Wort, Phrase usw.) zu beschreiben.

Ausgangspunkt der Überlegungen zu einer derartigen Methode, ist die Annahme, daß in einer natürlich-sprachlichen Äußerung das interne Timing optimal ist. Für eine experimentelle Untersuchung des Timings in einer nicht auf das Segmentelle gerichteten Methode bietet es sich an, gut verständliche parametrisierte Sprache (z.B. diphon-synthetisierte Äußerungen) zu benutzen, die äquivalent zu natürlicher Sprache ist. So gehen wir in dieser Arbeit zum einen von einer natürlich-sprachlichen Äußerung aus, zum anderen von deren diphon-konkateniertem Äquivalent.

## 2 Pilotstudie

Die Computerprogramme zur digitalen Sprachsignalverarbeitung am IPO sind so ausgelegt, daß sie eine Übertragung der Zeitstruktur einer natürlich-sprachlichen Äußerung auf das diphon-konkatenierte Äquivalent ermöglichen. Es werden die Dauerwerte der Segmente der natürlich-sprachlichen Äußerung übertragen, sodaß beide Sätze die gleiche Dauer bekommen (Siehe Abb.1 und 2).

Mittels der Resynthese können wir so das optimale Timing perzeptiv überprüfen. In einem weiteren Schritt können wir die resultierende Timingstruktur, die sich graphisch als Treppen (Siehe Abb.3) darstellt vereinfachen, so daß wir nur noch die perzeptiv relevanten Timingveränderungen innerhalb einer Äußerung erfassen. Im Idealfall erhalten wir so eine kontinuierliche Kontur des Timings.

### 2.1 Von Timing-Treppen zur Timing-Kontur

Die Erstellung einer Timing-Kontur erfolgt in folgenden Schritten:

- MRK: Segmentierung einer Äußerung und Transkription der Segmente.
- DS: Diphon-Konkatenierung und Synthese (ohne Dauerregeln) des entsprechenden Diphonfiles.
- MRK: Anpassung des Timings im Diphonfile an die originale Timingstruktur.
- CHT: Stilisierung der Timing-Kontur mit perzeptivem Feedback.

Hier mögen die Schritte im einzelnen erfolgen:

- Mrk: Mit dem Markierungsprogramm MRK ist es möglich, die Segmentation, oder besser ausgedrückt die Markierung von Segmenten, von analysierten Sprachfiles (im weiteren A/P-Files genannt) durchzuführen, wobei für den File das dazugehörige Oszillogramm unterstützend mit



aufgezeichnet wird. (A/P-Files sind analysierte Äußerungen natürlicher Sprache, die parametrisiert abgespeichert sind und die durch Resynthese wieder hörbar gemacht werden können (Die genaue Vorgehensweise entnehmen man Vogten (1983)). Die Markierung kann mit einer Genauigkeit von 1 Frame=10 msec durchgeführt werden.

- DS: Am Institut für Perzeptionsforschung (IPO) in Eindhoven wird Sprache mit Hilfe von Diphonen synthetisiert. Diphone werden gebildet, indem man aus natürlicher Sprache aus zwei aufeinanderfolgenden Phonemen die zweite Hälfte des ersten Phonems und die erste Hälfte des zweiten Phonems segmentiert. Dabei bleibt die Transition zwischen diesen beiden Phonemhälften erhalten und wird mit im Diphon abgespeichert, vgl. De Pijper (1988). Mit dem Programm DS werden die gleichen Sätze wie die natürlichen synthetisiert. Diese synthetischen Sätze werden in Diphonfiles abgespeichert. Im Pretest wurde das Dauermodul 3 benutzt, in dem nur das globale Satztempo geregelt wurde. Die Phonemdauerwerte wurden deshalb durch die inherenten Diphondauern bestimmt.

- MRK: Eine weitere Bearbeitung der A/P- und Diphonfiles wird auch noch innerhalb des Programms MRK durchgeführt, nämlich die Übertragung der internen Zeitstruktur des A/P-Satzes auf den Diphonsatz. Es werden die einzelnen Segmentlängen des Diphonsatzes an jenen des natürlichen Satzes angepaßt.

Beide Files werden segmentiert und mit einer Transkription versehen. In beiden Files muß sowohl die Anzahl der Segmente, als auch die Transkription übereinstimmen, wenn man den Diphonfile mit der internen Zeitstruktur des A/P-Files versehen will. Wir erstellen somit einen zeitangepaßten Satz, der tatsächlich genau die gleiche interne Zeitstruktur hat wie der Ausgangssatz. Wenn die beiden Sätze in ihrer zeitlichen Struktur gleich sind, kann der Diphonsatz mit der Original-F0 versehen werden (Programm TRQ). Wir erhalten also einen Diphonsatz, dessen Lautdauerwerte und dessen Grundfrequenz optimal sind.

- CHT: In dem Programm CHT werden die Dauerwerte von Phonemen graphisch als in Form von Strichen mit Labelwerten dargestellt. Das Programm ermöglicht die Manipulation von Labelwerten, und damit von Phonemdauern.

## 2.2 Ein Beispiel

Das Korpus der Sotscheksätze, das die Vorkommensfrequenz der deutschen Phoneme repräsentiert, bildet die Grundlage des Untersuchungsmaterials. An einem Satzbeispiel werden wir die weitere Vorgehensweise verdeutlichen.

Es wird der Originalsatz, "Heute ist schönes Frühlingswetter", mit seinem

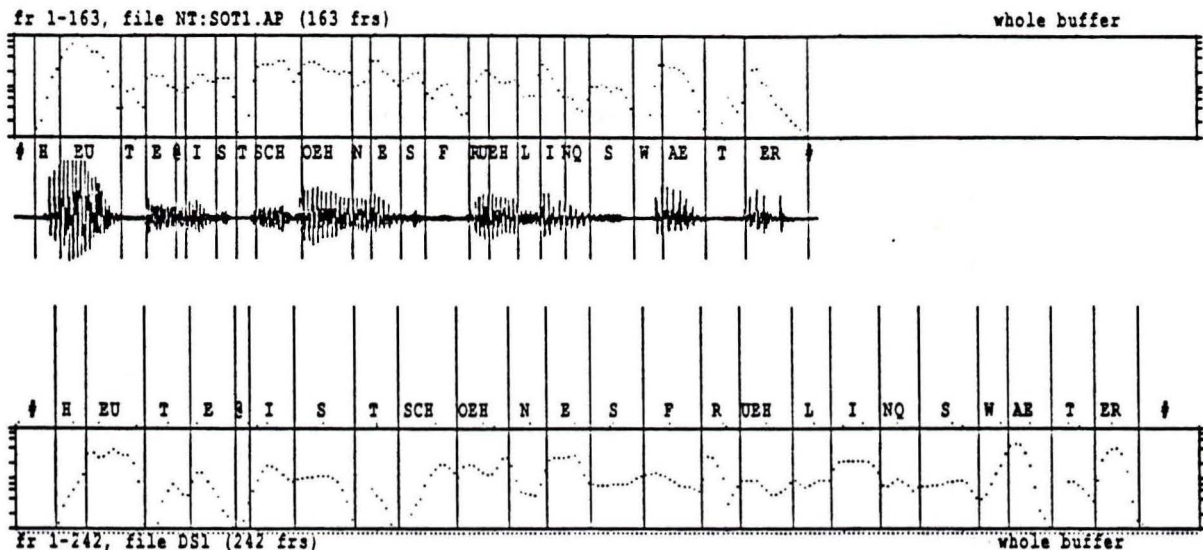


Abbildung 1: Natürlich-sprachlicher Originalsatz (oben) und passender Diphonsatz (unten). Vertikal sind die Segmentgrenzen angegeben.

Diphonäquivalent verglichen.

In Abb.1 sehen wir die Gegenüberstellung eines Originalsatzes (oben) und des mit dem Programm DS erstellten äquivalenten Diphonsatzes. Der Diphonsatz ist entscheidend länger als der Originalsatz, wir erwarten also, daß bei der Zeitanpassung die meisten Segmente verkürzt werden. In Abb.2 ist der Originalfile dem zeitangepaßten Diphonfile gegenübergestellt. Es ist ersichtlich, daß die Dauerwerte der definierten Segmente nun identisch sind.

Die Angabe des Verkürzungs- bzw. Verlängerungsfaktors für jedes Sprachlautsegment, im Programm CHT durch einen Labelwert dargestellt, nehmen wir als Grundlage, um eine Timing-Kontur zu entwerfen (Siehe Abb.3).

### 2.3 Pretest: Manuell erstellte Timing-Konturen

Es interessiert uns, ob dauerangepaßte synthetische Äußerungen für Hörer akzeptabel klingen. Neben den Versionen, in denen alle Frames pro Phonem das gleiche Label hatten, wurde deshalb eine Version erstellt, wobei die Labelwerte pro Frame in einem Phonem unterschiedlich waren (Ein hoher Labelwert bedeutet eine Verlängerung, ein niedriger eine



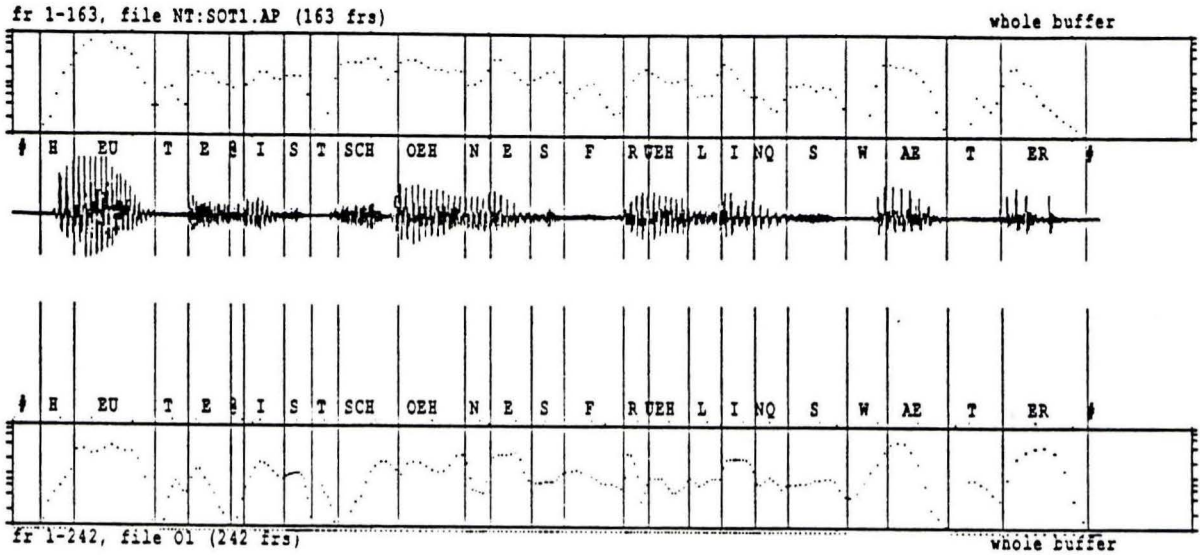


Abbildung 2: Natürlich-sprachlicher Originalsatz (oben) und zeitangepaßter Diphonsatz (unten). Vertikal sind die Segmentgrenzen angegeben.

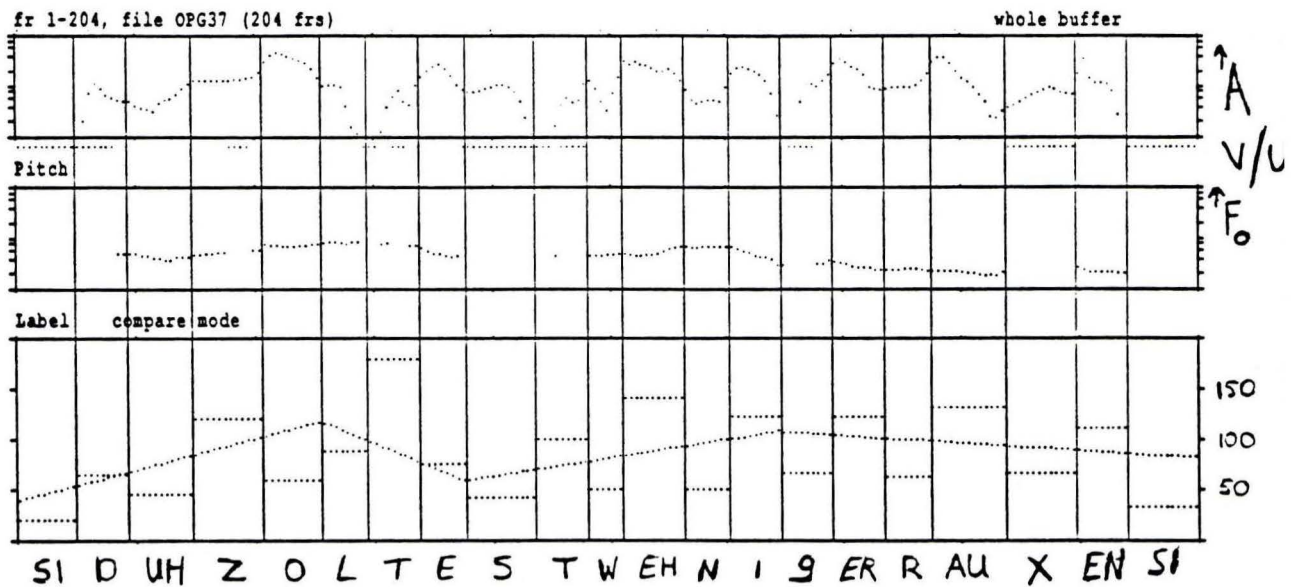


Abbildung 3: Aus den 'Treppen' wird die Timing-Kontur konstruiert. Vertikal sind die Segmentgrenzen angegeben.

Verkürzung). Hierdurch wird das Timing leicht verändert.  
Auf die in 2.1 beschriebene Weise haben wir daneben eine Anzahl von Timing-Konturen erstellt, die in einem Pretest evaluiert werden sollen.

### **2.3.1 Das Testmaterial**

Für den Test wurden 6 Sätze aus dem Sotschek-Korpus ausgewählt:

1. Löst doch die Fahrkarten am Schalter
2. Wir gehen auf den Bahnsteig
3. Du solltest weniger rauchen
4. Der Bahnhof liegt 7 Minuten entfernt
5. Da läuft der Zug ein
6. Die Station wird angesagt

Jeder dieser Sätze wurde in 4 Versionen synthetisiert (Siehe Abb.4):

DS: Diphonversion ohne Zeitanpassung

OPG: Diphonversion mit Originaltiming

CHT: Diphonversion mit leicht verändertem Originaltiming

TEST: Diphonversion mit Timing-Kontur

### **2.3.2 Durchführung**

Die Stimuli wurden paarweise zum Vergleich dargeboten. Auf dem Testbogen sollte die Frage beantwortet werden: "Welcher der beiden Sätze ist im Timing besser, der erste (A) oder der zweite (B)?"

Die Dauerwerte der verschiedenen Satzversionen, die miteinander verglichen wurden, stehen in Anhang A. Jede Version eines Satzes kam in 12 Stimuluspaaren vor (die 6 Paare pro Satz wurden in beiden möglichen Reihenfolgen in den Test aufgenommen).

Durchgeführt wurde der Test am Institut für Kommunikationsforschung und Phonetik (IKP) der Universität Bonn.

14 Versuchspersonen, Studenten und Mitarbeiter des IKP Bonn, von denen sich 3 mit Perzeptionsexperimenten auskannten, nahmen am Test teil.

### **2.3.3 Ergebnisse**

Da jede Version in 12 Stimuluspaaren vorkam und 14 Versuchspersonen am Test teilnahmen, konnte im Test maximal 1008 mal zugunsten einer der vier Versionen entschieden werden. Die Antworten wurden auf einen etwaigen Reihenfolgeeffekt überprüft. Die Überprüfung, wie die Antworten sich über A und B verteilen, ergab eine leichte, aber nicht-signifikante Verschiebung der Urteile zu A (2.2 von 36) (Siehe Tabelle 1).

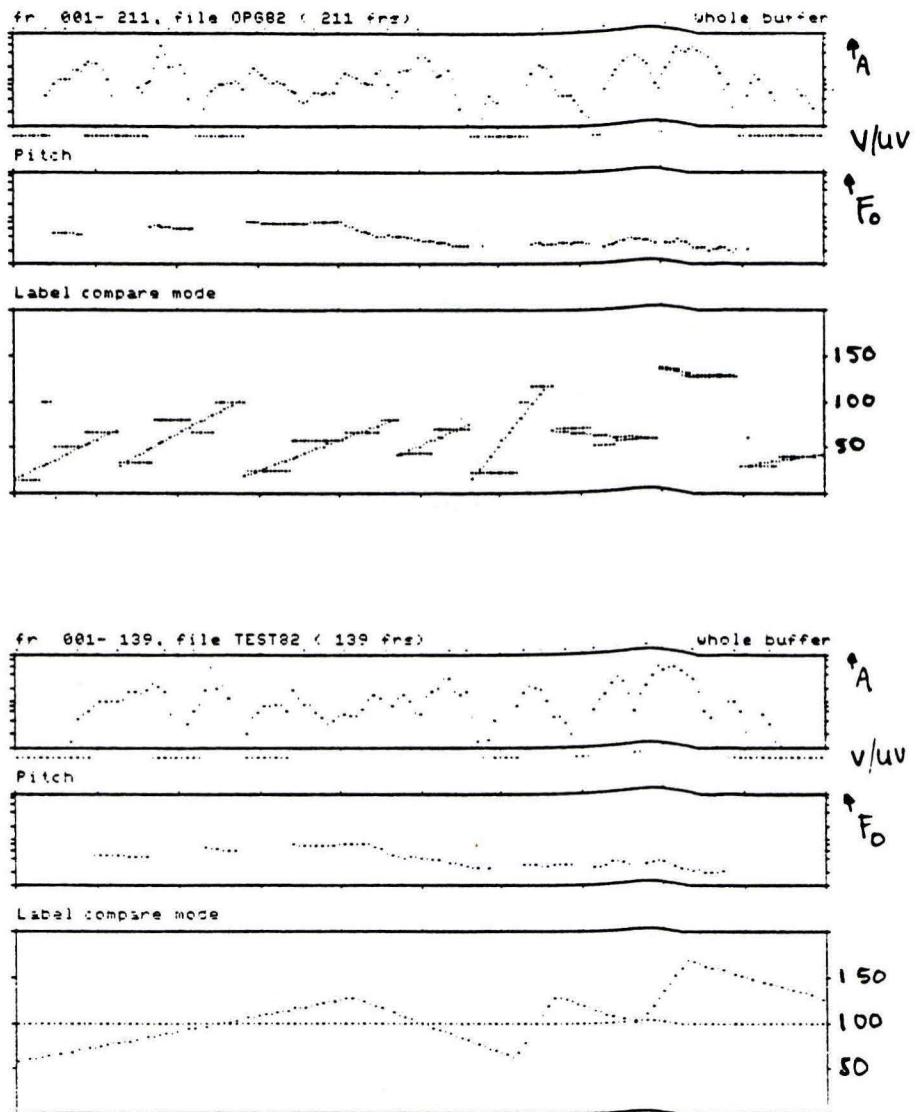


Abbildung 4: Die 4 Versionen eines Testsatzes. oben: Mit Originaltiming, und mit leicht verändertem Originaltiming, unten: Ohne Zeitanpassung, und mit Timing-Kontur.

*Tabelle 1: Verteilung der Antworten auf A und B der einzelnen VP.*

VP	A	B
1	38	34
2	34	38
3	34	38
4	46	26
5	40	32
6	39	33
7	32	40
8	37	35
9	45	27
10	35	37
11	34	38
12	40	32
13	45	27
14	36	36

Gemittelt entfielen von 72 auf

- A 38.2 (von 36)

- B 33.8 (von 36)



Da es keinen Reihenfolgeeffekt gab, werden die Ergebnisse von A und B im weiteren zusammengekommen.

In Tabelle 2 werden die Testergebnisse pro Version pro Satz, ermittelt über alle Versuchspersonen, aufgelistet.

*Tabelle 2: Entscheidungen zugunsten einer Version pro Satz (alle VP), absolute Zahlen, von insgesamt 1008 Urteilen.*

Satz	1	2	3	4	5	6	Summe
OPG	68	58	52	57	59	56	350
TEST	43	75	49	48	46	42	303
CHT	54	33	47	42	61	58	295
DS	3	2	20	21	2	12	60

Aus den Ergebnissen (Abb.5) wird deutlich, daß die Versuchspersonen insgesamt die Version OPG (d.h. die Diphonversion mit dem Originaltiming) bevorzugen.

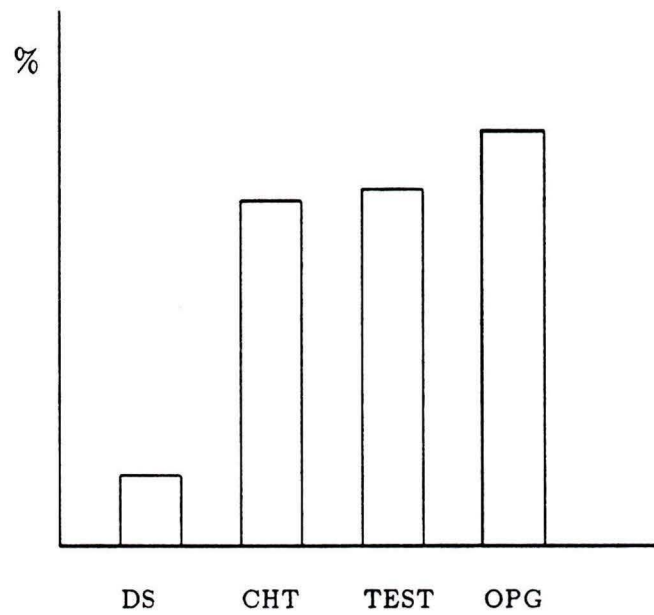
Die Version TEST (d.h. die Diphonversion mit Timing-Kontur) und die Version CHT (d.h. die Diphonversion mit leicht verändertem Originaltiming) wurden eindeutig der Version DS (d.h. die Diphonversion ohne Zeitanpassung) vorgezogen.

#### **2.3.4 Diskussion und Schlußfolgerungen**

Es interessierte uns, ob durch eine Timing-Kontur dauerangepaßte synthetische Äußerungen für Hörer akzeptabel klingen. Die Ergebnisse des Pretests zeigen, daß Stimuli mit Daueranpassung besser beurteilt werden, als Stimuli ohne Daueranpassung.

Die Ergebnisse bestätigen auch unsere Annahme, daß die Zeitstruktur einer natürlich-gesprochenen Äußerung als optimal betrachtet werden darf, denn die Stimuli mit der originalen Zeitstruktur werden gemittelt bevorzugt (Siehe Abb.5).

Es gibt jedoch einen Satz (Satz 2), bei dem die Version OPG nicht besser als die Version TEST abschneidet. Offensichtlich kommt die Timing-Kontur in diesem Fall den perzeptiven Relevanzen entgegen. Ein anderer Schluß wäre, daß die Diphone in diesem Satz nicht "sauber" klingen. Nochmaliges perzeptives Überprüfen der Stimuli dieses Satzes bestätigte unsere Vermutung und führte uns zur Überzeugung, daß das Timing am Satzende von Satz 2 in den Versionen OPG und CHT nicht korrekt war, nämlich zu schnell. In der Version TEST war dies zufällig korrigiert worden.



*Abbildung 5: Die Ergebnisse des Pretests im Überblick.*  
Von den 1008 Urteilen entfielen:

34,72 Prozent auf OPG  
30,06 Prozent auf TEST  
29,27 Prozent auf CHT  
5,95 Prozent auf DS

Die Ergebnisse zeigen, daß der Unterschied zwischen den Versionen CHT und TEST, (ausgenommen Satz 2), so klein ist, daß sich unsere Annahme der Möglichkeit nicht-segmentell orientierter Untersuchung des Satztimings als richtig erwiesen hat.

Die Arbeit mit den Stimuli des Pretests hat uns vermittelt, wie wir die neue Methode entwickeln können.

## 3 Die Methode

### 3.1 Timing-Stilisierung

Das in der Pilotstudie beschriebene Experiment hat den Ansatz dazu gegeben, die temporale Struktur eines Satzes kontinuierlich zu beschreiben. Aus der Arbeit mit den Diphonsätzen ist aber deutlich geworden, daß es notwendig ist, die Phonemdauerwerte zu standardisieren. Im Pretest wurde das Dauermodul 3 benutzt, in dem nur das globale Satztempo geregelt wurde. Die Phonemdauerwerte wurden deshalb durch die inherenten Diphondauern bestimmt. Wir erhielten durch die unterschiedlichen Phonemdauerwerte keinen Einblick mehr in die nur perzeptiv relevanten Timingaspekte, d.h. die 'Treppen' waren so hoch, daß keine passende Kontur gezogen werden kann (Siehe Abb.2). Es fehlte der Pilotstudie an der nötigen Systematik.

In der weiteren Arbeit werden daher die Phonemdauerwerte standardisiert. Das Standardisieren wird durchgeführt, indem die Phonemdauerwerte in der für die Synthese verwendeten Klatt-Tabelle angepaßt werden. In der Klatt-Tabelle (Siehe Anhang B) stehen die Phonemdauerwerte, die bei der Diphonsynthese erhalten werden sollen. Wir haben die Phoneme in Phonemklassen zusammengefaßt und jeweils einer Klasse einen festen Dauerwert zugeordnet.

Um das Stilisieren des Timings graphisch zu ermöglichen, haben wir außerdem die Framedauerwerte auf 10 ms standardisiert.

Im Vergleich zu der in der Pilotstudie beschriebenen Vorgehensweise haben wir einige Änderungen durchgeführt.

Die Erstellung verläuft in den folgenden Schritten (Siehe Abb.6):

- MRK: Segmentierung einer Äußerung und Transkription der Segmente.
- DS: Diphon-Konkatenierung und Synthese des entsprechenden Diphonfiles. (Ohne Dauerregeln, feste Phonemdauerwerte in Tabelle)
- CFR: Korrektur der Framedauerwerte auf 10 ms.



- MRK: Anpassung des Timings im Diphonfile an die originale Timingstruktur.
- CHT: Kontrolle der Timingstruktur.
- Anpassung der Tabellenwerte, damit eine möglichst glatte Timing-Kontur bekommen wird (Die Phoneme werden soviel wie möglich in Phonemklassen zusammengefaßt).
- TRQ: Übertragen der Originalintonation.
- CHT: Stilisierung der Timing-Kontur mit perzeptivem Feedback.
- Perzeptive Evaluierung der Stilisierungen.

### 3.1.1 Das Sprachmaterial

Als Sprachmaterial haben wir die Sotscheksätze verwendet. Diese Sätze wurden von einem deutschen Nachrichtensprecher vorgelesen, danach digitalisiert und segmentiert/transkribiert. Zuerst haben wir die phonemische Transkription der 100 originalen Sotscheksätze kontrolliert, was eine Vergleichung der Synthese- und der Originalversion eines jeden Satzes ermöglichte. Diese phonemische Transkription ist Input für die Diphonsynthese. Der verwendete Graphem-Phonem-Konverter ist Nummer 2, in den der Text in Phonemschrift eingegeben wird. Die Dauerregeln des Dauermodul 2 (sogenannte Klatt-Regeln) sind so eingestellt, daß die Phonemdauerwerte, die in einer Klatt-Tabelle aufgelistet sind, nicht geändert werden.

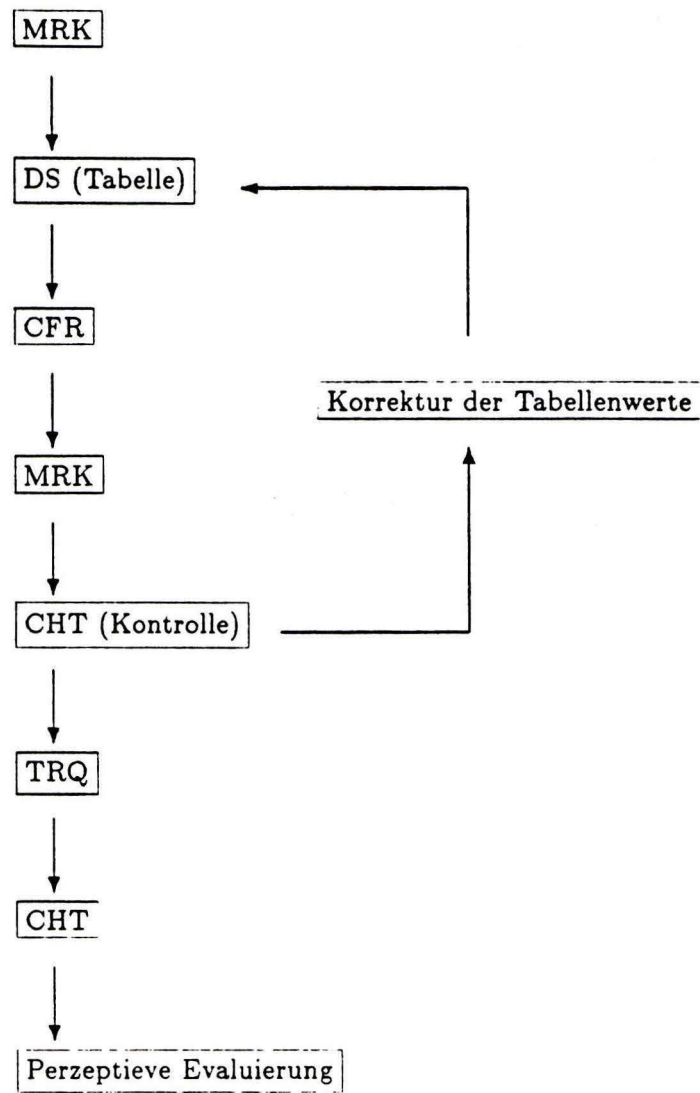
Nach der Synthese wird das Timing und die Intonation der Originalsätze auf die DS-Sätze übertragen, so daß wir auf diese Weise synthetische Sätze mit einem natürlichem Timing und einer natürlichen Intonation bekommen.

### 3.1.2 Phonemdauern

Mit der Diphonsynthese bekommen die Phoneme einen Dauerwert, der in der Klatt-Tabelle aufgelistet ist. Die Phonemdauerwerte der Tabelle sind so gewählt worden, daß für alle Sätze eine möglichst glatte Timingstruktur ohne große Sprünge entsteht, wenn das Originaltiming auf die Diphonsätze übertragen wird (Siehe Abb.8). In der Tabelle in Anhang B ist das Ergebnis der Anpassung der Segmentdauerwerte (nachdem einige Male der Zyklus (Abb.6) durchlaufen ist) aufgelistet.

Nach der Diphonsynthese wird mit dem Programm CFR der Framedauerwert aller Frames auf 10 ms standardisiert (Abb.7). Im zweiten (H)-Phonem werden also die 5 Frames von 12 ms (mit Label 120) durch 6





*Abbildung 6: Die Erstellung der Stilisierungen.*

Frames von 10 ms (mit Label 100) ersetzt. Alle Phoneme bekommen auf diese Weise einen Labelwert von 100: die Timingstruktur eines Diphonsatzes nach der Synthese und nach CFR ist also graphisch eine gerade Linie.

### 3.1.3 Das Übertragen der Timingstruktur

Die Timingstruktur des Originalsatzes besteht aus den verschiedenen Phonemdauerwerten. Das Übertragen der Timingstruktur beinhaltet, daß die Phoneme im synthetischen Satz den gleichen Dauerwert bekommen wie die Phoneme im Originalsatz. (Vergleiche 2.2: 'Ein Beispiel'). Mit dem LVS-programm MRK wird das Timing der Originalsätze auf die synthetisierten Sätze (DS-Sätze) übertragen: sie haben dann beide das gleiche Satztiming und also auch die gleiche Dauer.

Der Dauerwert jedes Phonems wird graphisch als ein Strich mit einem Labelwert dargestellt. Das Übertragen des Timings führt dazu, daß Phoneme des Diphonsatzes in einigen Fällen verlängert und in anderen Fällen verkürzt werden: eine Verkürzung ergibt einen Labelwert unter 100, eine Verlängerung ergibt einen Labelwert über 100. Das Timing wird deshalb graphisch als eine Treppchenstruktur dargestellt (Siehe Abb.8).

Die DS-Sätze waren immer länger als die Originalsätze (Siehe Tabelle im Anhang D). Die meisten Phoneme wurden deshalb verkürzt (manche sogar bis auf 5% des Dauerwertes, der als Tabellenwert angegeben war) und Verlängerungen (bis zu 160%) kamen nur selten vor.

### 3.1.4 Die Stilisierungen

Die Arbeitsweise ist analog zu der, wie sie am IPO bei der Intonationsforschung verwendet worden ist. Dabei wurden nur die perzeptiv relevanten Schwankungen der natürlichen Intonationskurve berücksichtigt. De Pijper (1983) hat die so stilisierten Intonationskonturen 'Close Copies' genannt.

Für die Stilisierungen haben wir die synthetischen Sätze mit dem Originaltiming als Ausgangspunkt gewählt. Die Stilisierung soll sich aus einer minimalen Anzahl gerader Linien zusammensetzen. Außerdem sollen die Linien über den ganzen Satz einen ununterbrochenen Verlauf aufweisen und somit eine Timing-Kontur bilden (Siehe Abb.9).

Die Timing-Kontur der stilisierten Sätze ist eine Vereinfachung der Timingstruktur. Die Sätze mit dem stilisierten Timing und die Sätze mit dem Originaltiming sollen perzeptiv identisch sein.

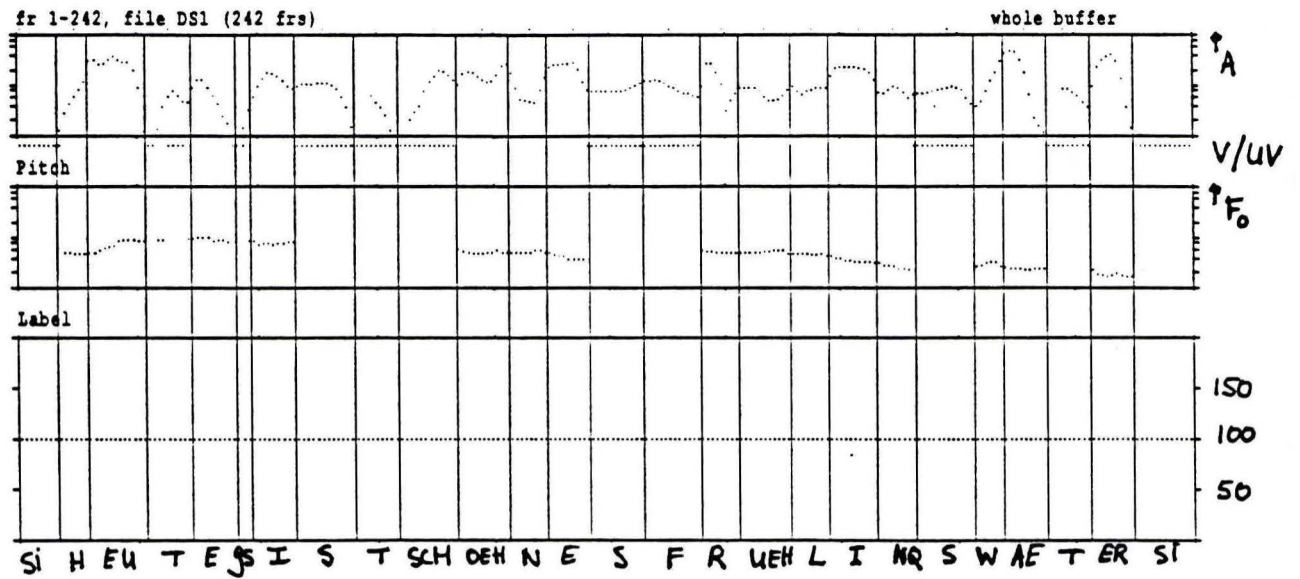
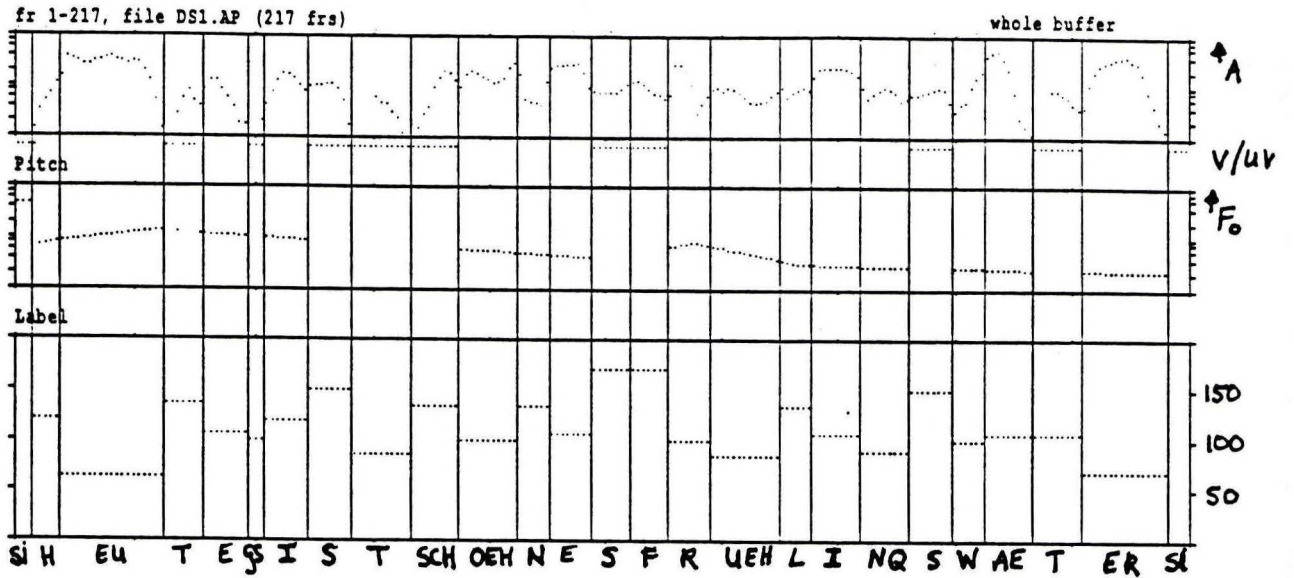


Abbildung 7: Timingstruktur vor (oben) und nach CFR (unten). Vertikal sind die Segmentgrenzen angegeben. Jeder Punkt im Label-Feld repräsentiert ein Frame.

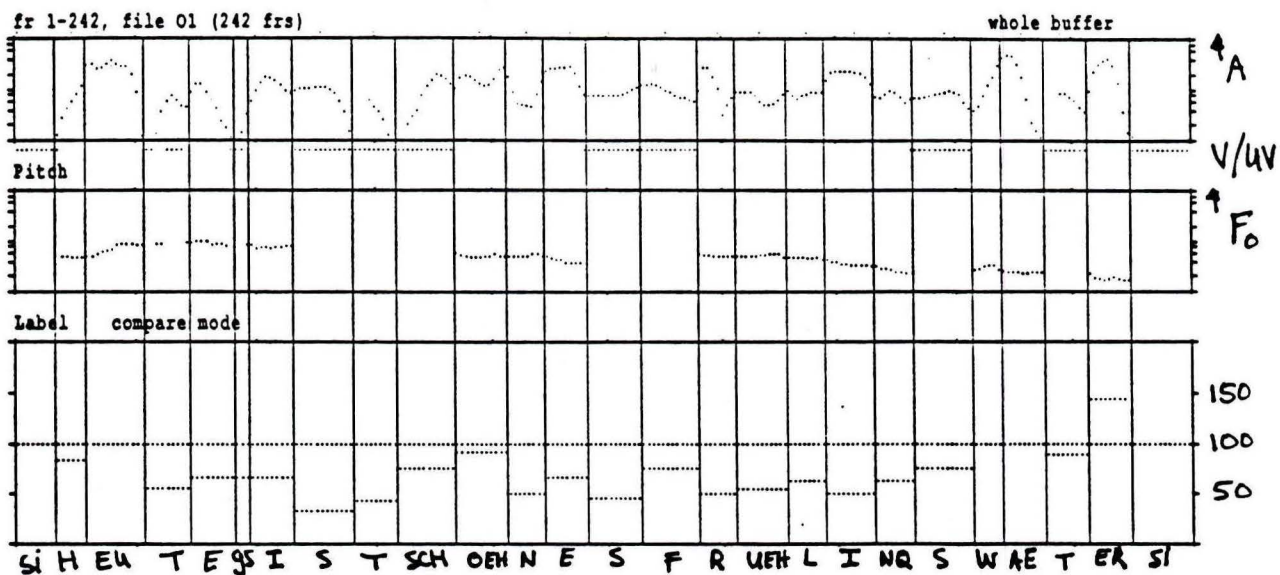


Abbildung 8: Timingstruktur des Diphonsatzes mit dem Originaltiming. Vertikal sind die Segmentgrenzen angegeben. Zum Vergleich ist die Label-100-Linie mit abgebildet.

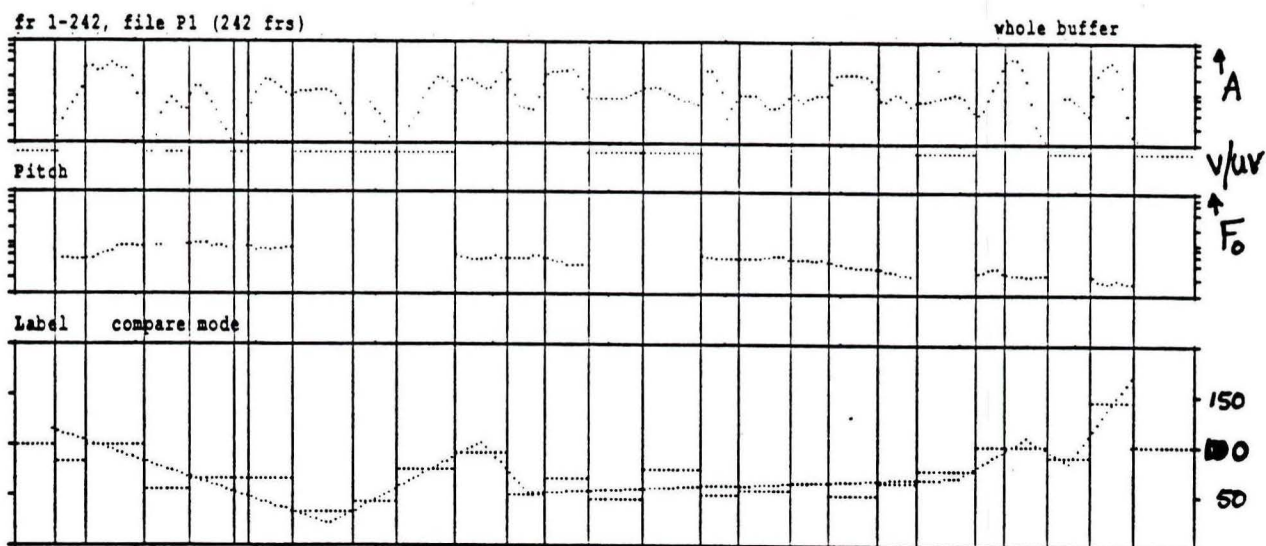


Abbildung 9: Timing-Kontur nach der Stilisierung. Die Segmentgrenzen sind als vertikale Linien angegeben.



Der Arbeitsvorgang beim Stilisieren in CHT ist folgender: zuerst werden die Phonemlabel soviel wie möglich durch gerade Linien verbunden. Beim Stilisieren kann das Ergebnis interaktiv perceptiv mit dem Satz mit dem Originaltiming und dem resynthetisierten Originalsatz verglichen werden. Unterschiede zu diesen Sätzen werden durch Anpassung der Stilisierung auf das Geringste reduziert.

## **3.2 Perzeptionstest: Timing-Stilisierungen**

### **3.2.1 Das Testmaterial**

Nach dem Stilisieren war jeder Satz in zwei Versionen vorhanden: einmal eine Version mit dem Originaltiming und eine Version mit dem stilisierten Timing. Die Version mit dem Originaltiming werden wir weiterhin 'O-Sätze' nennen, die Version mit dem stilisierten Timing 'P-Sätze'.

Diese beiden Versionen haben Versuchspersonen paarweise perceptiv verglichen. Die Frage lautete: sind die beiden Versionen im Timing gleich oder ungleich? (Siehe Testantwortbogen im Anhang F).

Damit die Versuchspersonen nicht nur gleiche oder fast gleiche Sätze vergleichen mußten, haben wir von jedem O-Satz auch einen Ableitersatz (Distractor) gewählt. In diesem Ableiter- (oder 'A'-)Satz war die Timing-Kontur so hergestellt, daß es perceptiv tatsächlich einen Unterschied zu den O- und P-Sätzen gab. Die Timingstruktur wurde in diesem Fall durch eine gerade Linie ersetzt, die der Timing-Kontur des P-Satzes aber soviel wie möglich folgte (Abb.10 und Anhang E).

Auf diese Weise haben wir also drei Versionen eines Satzes bekommen:

- O-Satz: Diphonsatz mit natürlichem Timing,
- P-Satz: Diphonsatz mit perceptiv identisch stilisiertem Timing, und
- A-Satz: Diphonsatz mit perceptiv unterschiedlich stilisiertem Timing.

### **3.2.2 Durchführung des Testes**

Das Testband:

Von den 100 Sotscheksätzen haben wir etwa vierzig Sätze stilisiert. Aus diesen vierzig Stilisierungen haben wir die fünfzehn besten P-Sätze ausgesucht, um sie von Versuchspersonen perceptiv mit den dazugehörigen O- und A-Sätzen vergleichen zu lassen (Anhang C). Der Vergleich sollte paarweise geschehen.

Es wurden pro Satz 8 Testpaare hergestellt: Die Kombinationen waren O-O, P-P, O-P, P-O, O-A, A-O, P-A und A-P. Da fünfzehn Sätze getestet

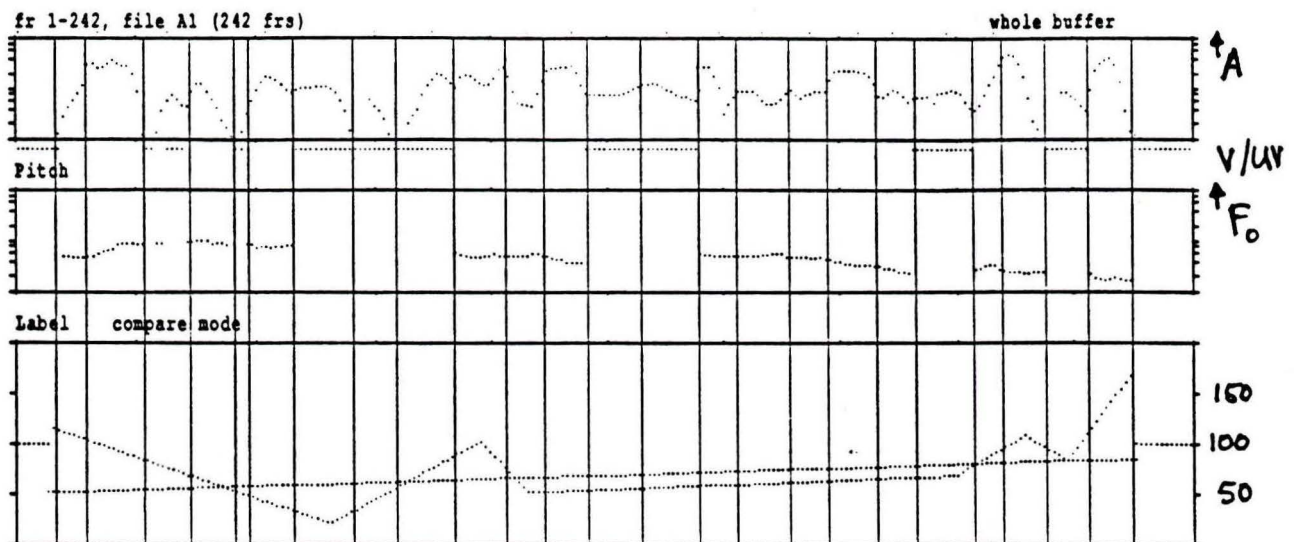


Abbildung 10: Timing-Kontur der Ableiterversion. Vertikal sind die Segmentgrenzen angegeben.

wurden, gab es im ganzen also 120 (= 8 x 15) Testpaare. Die Testpaare wurden in randomisierter Reihenfolge auf eine Tonbandkassette aufgenommen. Die Sätze dauerten je etwa 2 Sekunden. Zwischen den Sätzen innerhalb eines Paares gab es eine Pause von 350 ms. Ein Testpaar dauerte also etwa siebeneinhalb Sekunden. Zwischen den Testpaaren gab es eine Pause von 3500 ms, in der die Versuchspersonen das Item im Testantwortbogen ankreuzen sollten. Der Test dauerte 18 Minuten. Vor Anfang des eigentlichen Testes wurde den Versuchspersonen einige Minuten synthetische Sprache vorgespielt, damit sie sich an die Diphonsprache gewöhnen konnten. In diesen Vorspann waren alle Testsätze einmal (als O-P-Paar) mit aufgenommen.

Testdatum, Testort, Versuchspersonen:

Den Test hat der zweite Autor am Institut für Kommunikationsforschung und Phonetik (IKP) der Universität Bonn durchgeführt. Die Versuchspersonen waren Studenten und Mitarbeiter des Instituts und hatten als Muttersprache Deutsch. Es haben sich 11 Versuchspersonen (VP) am Test beteiligt.

Tabelle 3: Die Scores der VP.

'Gleich'-Scores fuer die verschiedenen VP

VPnr.	Identische:		Testsätze:		Ableiter:			
	O-O	P-P	O-P	P-O	A-O	O-A	A-P	P-A
1	12	10	8	12	11	8	10	9
2	11	10	11	10	12	6	12	4
3	12	11	10	12	11	3	10	3
4	11	14	11	12	12	1	12	3
5	10	10	12	8	3	1	4	1
6	14	15	15	15	12	8	13	4
7	14	11	12	12	10	6	11	6
8	12	7	12	11	7	4	8	1
9	14	15	13	14	10	5	13	3
10	11	14	13	12	7	2	8	1
11	13	15	13	10	12	1	10	1
Gesamt:	134	132	130	128	107	45	111	37

### 3.2.3 Ergebnisse

Die Summe der Scores beträgt 1320 (= 11 x 120). Davon waren 824 'Gleich'-Scores (62 % aller Scores).

In den Tabellen 3 und 4 stehen die 'Gleich'-Scores der VP und die 'Gleich'-Scores für die Sätze aufgelistet.

Die 11 Vp sollten 2 mal 15 identische Sätze beurteilen: 15 mal O-O und 15 mal P-P. Das Ergebnis war, daß von den 330 Scores 264 mal 'gleich' geurteilt wurde (80 %): O-O 133 mal, und P-P 131 mal. Das heißt, daß 20 Prozent der identischen Stimuli noch als ungleich beurteilt wurden.

Die Scores für die anderen Kombinationen waren (von 330 Scores): O-P/P-O 258 mal 'gleich' (78 %), A-P/P-A 165 mal 'gleich' (50 %), und A-O/O-A 156 mal 'gleich' (47 %) (Abb.11).

Mit Hilfe eines Zeichentests haben wir getestet, ob es einen signifikanten Effekt der Stilisierung gab. Die Signifikanzgrenze haben wir auf 5% gelegt. Das Ergebnis war, daß es sowohl pro VP als pro Satz keinen signifikanten Effekt der Stilisierung gab.

- VP: N=9 (zwei 'ties'), 7 mal +, 2 mal - :  $p = .09$

- Sätze: N=13 (auch zwei 'ties'), 7 mal +, 6 mal - :  $p = .5$

Eine weitere Analyse der VP-Ergebnisse mit Hilfe eines 'Wilcoxon matched pairs signed ranks test' ergab auch einen nicht-signifikanten Stilisierungseffekt ( $W = 10.5$ , Grenzwert für Signifikanz ist 6).

In Abb.11 kann man sehen, daß die Reihenfolge innerhalb der getesteten



Tabelle 4: Die Scores für die Sätze.

'Gleich'-Scores fuer die verschiedenen Saetze

Satznr.	Identische:		Testsaetze:		Ableiter:			
	O-O	P-P	O-P	P-O	A-O	O-A	A-P	P-A
1	10	8	11	11	6	6	7	1
2	10	8	9	9	10	2	8	0
3	7	10	9	8	7	2	9	3
4	7	8	10	9	10	5	10	5
5	9	7	10	9	7	0	7	3
6	11	7	5	6	6	0	4	1
7	10	9	10	8	3	1	5	0
8	8	9	8	7	7	3	7	3
9	10	11	10	8	7	4	10	4
10	7	9	7	10	4	1	4	1
11	5	9	8	3	6	4	8	1
12	10	10	5	10	11	9	8	9
13	10	9	8	9	9	5	9	4
14	10	9	10	11	7	1	6	2
15	10	9	10	10	7	2	9	0
gesamt:	134	132	130	128	107	45	111	37

Tabelle 5: Die Verteilung der 'Gleich'-Scores. Vertikal: erster Satz im Vergleichspaar, horizontal: zweiter Satz im Vergleichspaar.

	O	P	A
O	134	130	45
P	128	132	37
A	107	111	



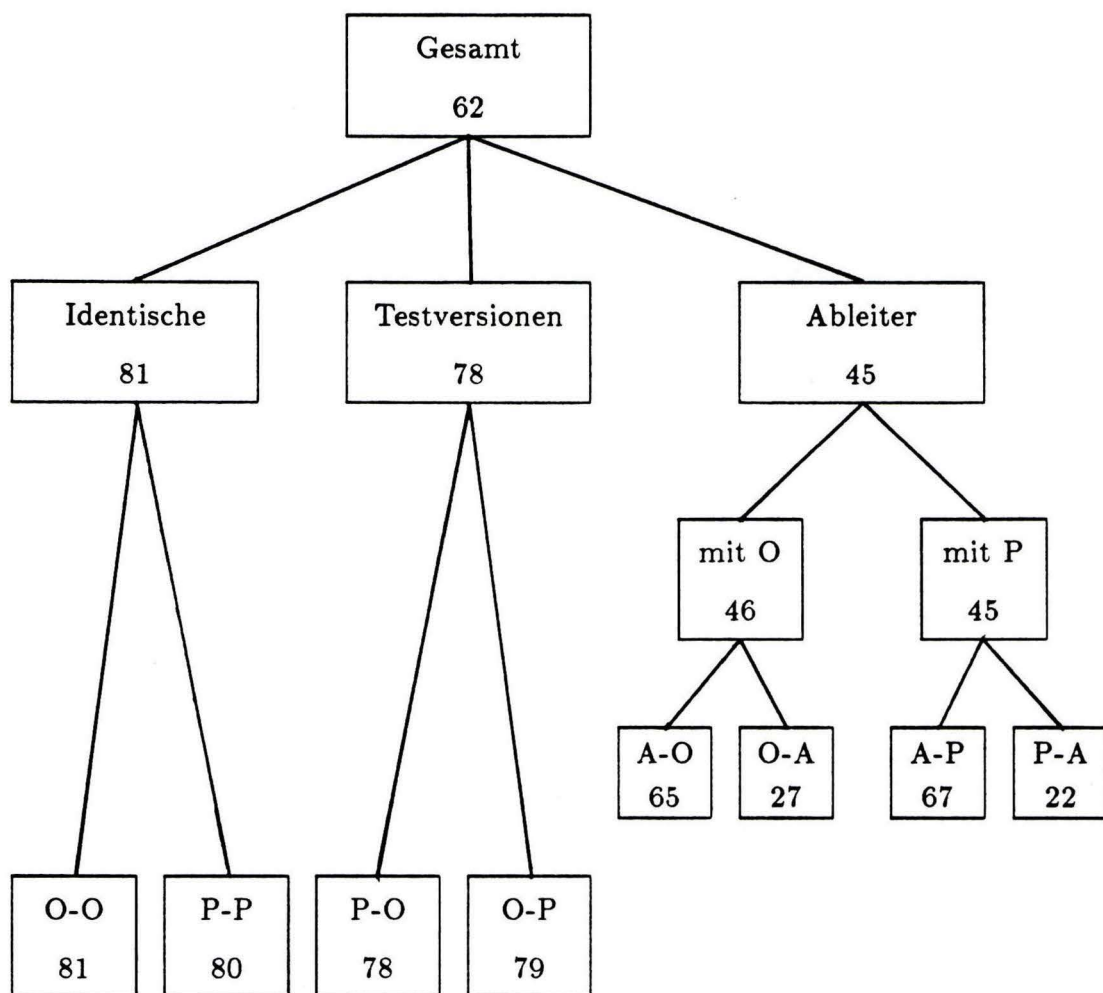


Abbildung 11: Die Scoreverteilung der 'Gleich'-Antworten in Prozent.

Paare (die O-P- und P-O-Paare) keinen Effekt hatte. Bei den Ableitersätzen haben wir einen großen Unterschied zwischen den beiden Scores (A-P: 111, P-A: 37 und A-O: 107, O-A: 45) gefunden. Eine Analyse mit demselben Wilcoxon-Test ergab einen sehr signifikanten Reihenfolge-Effekt ( $p < .001$ ).

### 3.2.4 Diskussion und Schlußfolgerung

Der Vergleich von timingstilisierten Sätzen und Sätzen mit Originaltiming zeigt, daß die VP zwischen diesen beiden Versionen nicht signifikant öfter einen Unterschied hören, als zwischen zwei identischen Sätzen. Die Ableitersätze werden von den VP dahingegen signifikant öfter von den P- und O-Sätzen verschieden beurteilt.

Es zeigte sich, daß die Reihenfolge innerhalb eines Testpaares einen signifikanten Effekt hatte. Wenn der stilisierte oder der Originalsatz zuerst kam, hörten die VP den Unterschied öfter als wenn der A-Satz zuerst kam. Es ist leider nicht deutlich, woher dieser doch sehr auffallende Reihenfolge-Effekt stammt. Eine Möglichkeit ist, daß die VP irgendeine 'interne Repräsentation' des richtigen Timings haben, mit der jeder gehörte Satz verglichen wird. Ein 'verkehrtes' Timing gelangt nicht zur internen Repräsentation wodurch ein Vergleich mit einem Satz mit richtigem Timing nicht zustande kommt. Wenn ein Satz mit falschem Timing als zweiter im Paar angeboten wird, wird der Vergleich mit dem Satz mit richtigem Timing im Kurzzeitgedächtnis ermöglicht. (Diese Unsicherheit der VP dürfte auch die Erklärung dafür sein, daß noch 20 Prozent der identischen Paare als 'ungleich' beurteilt wurden.)

Daß es beim Vergleich der O- mit den P-Sätzen nicht diesen Reihenfolge-Effekt gab, deutet darauf, daß die Stilisierungen genügend mit der 'internen Repräsentation' übereinstimmten, um als 'richtig' erkannt zu werden. Aus den Ergebnissen dieses Testes kann man also folgern, daß es möglich ist, perzeptiv-identische Timing-Konturen zu entwerfen.

## 4 Diskussion

Die Pilotstudie und das Experiment haben gezeigt, daß synthetische Sätze mit dem Timing ihrer Originalversionen von Versuchspersonen stark bevorzugt werden, im Vergleich zu synthetischen Sätzen ohne Daueranpassung. Eine gute Sprachsynthese zeichnet sich dadurch aus, daß sie möglichst natürlich klingt. Ziel ist es deshalb, synthetische Sätze mittels eines Regelsystems mit einem möglichst natürlich klingenden Timing zu versehen.

Bevor man zu einem Regelsystem gelangen kann, ist es notwendig, die

Regelmäßigkeiten im natürlichem Satztiming zu kennen. Die Voraussetzung dabei ist, daß die Annahme stimmt, daß das Timing in Originalsätzen optimal natürlich ist. Die perzeptive Evaluierung hat gezeigt, daß sowohl die Annahme des optimalen Timings in Originalen stimmt, als auch, daß es möglich ist, das Timing von Originalsätzen als Close Copies zu stilisieren. Durch die Methode des Stilisierens können die Regelmäßigkeiten des Timing ermittelt werden. Die Regelmäßigkeiten lassen sich dann in einem Timing-Modell zusammenfassen.

Wie schon in der Einleitung erwähnt, wird der Dauerwert der einzelnen Phoneme durch die Stellung des Phonems im Satz, durch den Kontext in dem das Phonem steht, durch den Satzakzent und durch das Satztempo bestimmt. Dies wird bestätigt durch die folgenden Regelmäßigkeiten, die dem zweiten Autor beim Stilisieren aufgefallen sind:

- Am Satzende werden Segmente verlängert;
- An Stellen, wo es im Satz eine Intonationsbewegung gibt, kommt oft eine Verlängerung von Segmenten vor;
- Vor einer Satzpause oder einer prosodischen Grenze gibt es eine Verlängerung von Segmenten;
- Nach jeder Verlängerung gibt es auch eine Verkürzung von Segmenten (manchmal sogar am Ende eines Satzes);
- Lange Vokale und Diphtonge werden am meisten verlängert;
- Am Ende eines Satzes wird entweder der letzte Vokal verlängert, oder im Falle eines Kurzvokals der darauffolgende sonorante Konsonant.

Die gefundenen Regelmäßigkeiten lassen schließen, daß Segmentgruppen nicht als linguistische Einheiten wie Silbe, Wort oder Phrase zu beschreiben sind. Sie zeigen aber am meisten die Struktur von Silben auf.

Betrachten wir die Regelmäßigkeiten, die uns bei der Arbeit mit den Timing-Konturen aufgefallen sind, so zeigt sich eine enge Verknüpfung von Intonation und Timing. Im vorliegenden Sprachsynthesystem des IPO sollte daher erwägt werden, ob man nicht die intonationssteuernden und die timingsteuernden Module aneinander koppelt, indem man das Timing in Abhängigkeit von dem Intonationmarkierungen errechnen läßt.

Mit der hier vorgestellten Methode lassen sich auch auf analoge Weise die temporalen Strukturen anderer Sprachen ermitteln. Wir schlagen vor, für

das am IPO verfügbare multi-linguale Synthesystem eine einheitliche Timing-Handhabung zu entwickeln, die auf dem Timing-Modell der jeweiligen Sprache beruht.



## 5 Literaturverzeichnis

- ELSEENDOORN, B. A. G., 1984, Production and perception of English vowel duration by Dutch speakers of English. IN: International congress of phonetic sciences, 10th., 1983. Ed. by M. P. R. van den Broecke and A. Cohen. Dordrecht: Foris, 673-676.
- FISCHER-JØRGENSEN, E., 1964, Sound duration and place of articulation IN: Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung Bd. 17, 175-207.
- KLATT, D. H., 1987, Review of text-to-speech conversion for English. IN: J.Acoust.Soc.Am., Bd. 82, 737-793.
- NOOTEBOOM, S. G., 1972, Production and perception of vowel duration. Dissertation Universit t Utrecht.
- NOOTEBOOM, S. G., 1973, The perceptual reality of some prosodic durations. IN: Journal of phonetics, Bd. 1, 25-45.
- NOOTEBOOM, S. G., en COHEN, A., 1984, Spreken en verstaan: een nieuwe inleiding tot de experimentele fonetiek; Assen: Van Gorcum.
- PETERSON, G. E. AND LEHISTE, I., 1960, Duration of syllable nuclei in English. IN: J.Acoust.Soc.Am.,Bd. 32, 693-703.
- PIJPER, J. R. de., 1983, Modelling British English intonation: an analysis by resynthesis of British english intonation. Dissertation Universit t Utrecht.
- PIJPER, J. R. de., 1988, Van tekst naar spraak met behulp van difoonsynthese. IN: Tijdschr. voor Taal- en tekstwetenschap, Bd. 3, 225-237.
- VOGTEN, L. L. M., 1983, Analyse, zuinige codering en resynthese van spraakgeluid. Dissertation Technische Universit t Eindhoven.

## 6 Anhang A: Die Satzdauerwerte der 4 Versionen im Pretest

Satznr.	Version	Sekunden
1	DS	1.46
1	OPG	1.43
1	CHT	1.51
1	TEST	1.56
2	DS	2.05
2	OPG	2.05
2	CHT	1.95
2	TEST	2.12
3	DS	1.77
3	OPG	1.79
3	CHT	1.72
3	TEST	2.02
4	DS	1.25
4	OPG	1.28
4	CHT	1.31
4	TEST	1.59
5	DS	1.08
5	OPG	1.10
5	CHT	1.13
5	TEST	1.10
6	DS	1.41
6	OPG	1.41
6	CHT	1.31
6	TEST	1.51

## 7 Anhang B: Die Klatt-Tabelle

Phonem	Min	Max	Phonembeispiel
I	95	95	Ich
IE	110	110	Ihre
AE	95	95	Elf
EH	110	110	Eben
E	85	85	stunde
A	95	95	April
AH	110	110	Aber
O	95	95	Onkel
OH	110	110	Oder
U	95	95	Und
UH	110	110	Uhr
UE	95	95	bUero
UEH	110	110	UEbel
OE	95	95	Oeffentlich
OEH	110	110	OEfen
AI	120	120	Elfrig
AU	130	130	AUch
EU	120	120	EURE
P	100	100	Pater
T	90	90	TaT
K	100	100	Kater
B	90	90	Bad
D	90	90	Da
G	90	90	Gabel
F	120	120	Fahrrad
S	120	120	daS
X	120	120	aCH
CH	120	120	iCH
SCH	120	120	SCHon
Z	120	120	Sehen
M	80	80	Mutter
N	80	80	Natur
NQ	80	80	juNQe
W	60	60	Wahr
Y	100	100	Ja
H	60	60	Hoffen
L	80	80	Lang
R	80	80	Rot
SI	250	250	(silence)
GS	30	30	(glottal stop)
EL	90	90	edEL
EM	90	90	jedEM
EN	90	90	jedEN
ER	90	90	jedER

## 8 Anhang C: Die verwendeten Sotscheck-sätze im Perzeptionstest

Die verwendeten Sätze:

1. Heute ist schönes Frühlingswetter.
2. Über die Felder weht ein Wind.
3. Gestern stürmte es noch.
4. Es ist acht Uhr morgens.
5. Wer möchte keinen Kuchen?
6. Wer möchte noch Milch?
7. Muß der Zucker nicht dort drüben stehen?
8. Hier gibt es Konserven.
9. Ich müßte lesen und rechnen.
10. Was macht denn dein verstauchter Fuß?
11. Ich spüre ihn nicht mehr.
12. Da möchte ich gerne mit.
13. Am Zaun steht eine Regentonne.
14. Ist es nicht Zeit zum Aufbruch?
15. Das war jetzt aber ein schöner Tag.

Die Sätze in Phonemschreibweise:

H'EUTE GSIST SCHOEHNES FR'UEHLINQSWAETER.  
GSAES GSIST GS'AXT GSUHER M'ORGENS.  
D'AS WAHER YAETST GSAHBER GSAIN SCH'OEHNER TAHK.  
WEHER MOECHTE K'AINEN K'UHXEN?  
WEHER MOECHTE NOX M'ILCH?  
MUS DEHER TS'UKER NICHT DORT DR'UEHBEN SCHTEHEN?  
HIEER GIEPT GSAES KONZ'AERW EN.  
GSUEHBER DIE F'AELDER WEHT GSAIN W'INT.  
GSICH MUESTE L'EHZEN GSUNT R'AECHNEN.  
WAS M'AXT DAEN DAIN FAEERSCHT'AUXTER FUHS?  
GSICH SCHP'UEHRE GSIEN NICHT MEHER.  
DAH MOECHTE GSICH G'AEERNE MIT.  
GAESTERN SCHT'UERMTE GSAES NOX.  
GSAM TS'AUN SCHTEHT GSAIN R'EHGENTONE.  
F'AHTER HAT DEHN T'ISCH GEDAECT.

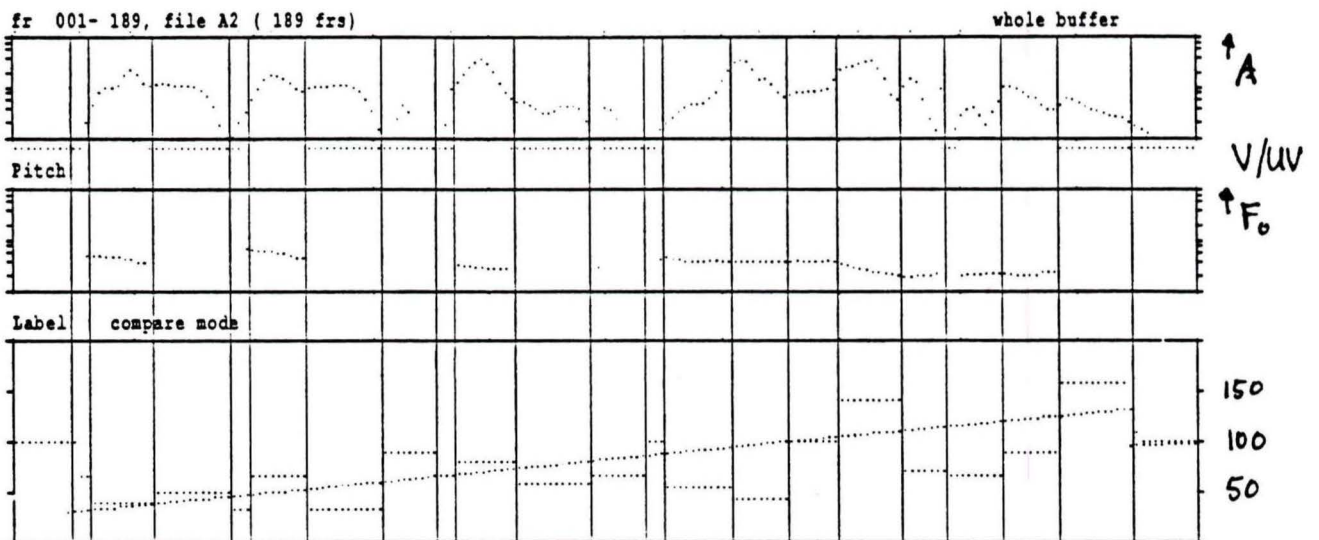
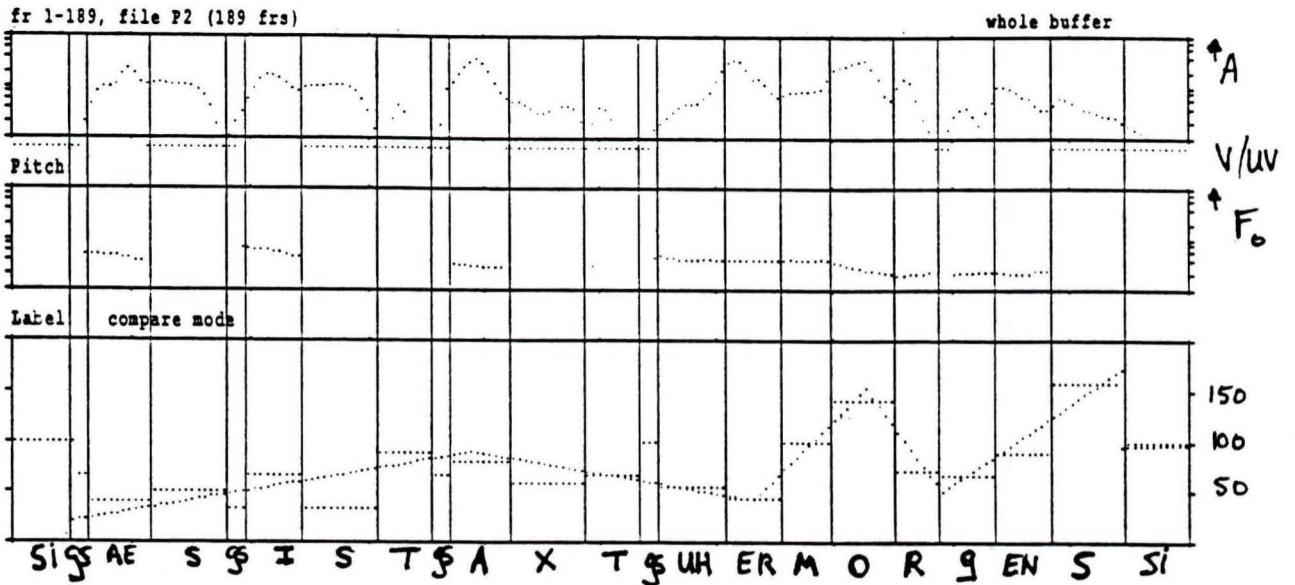


## 9 Anhang D: Die Satzdauerwerte der Versionen im Perzeptionstest

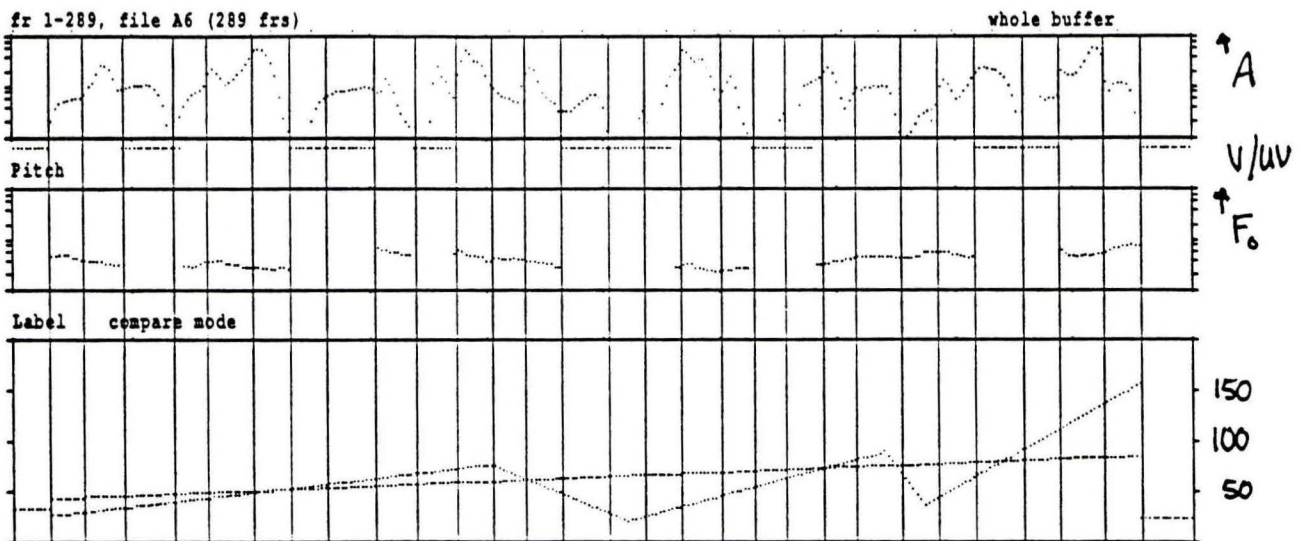
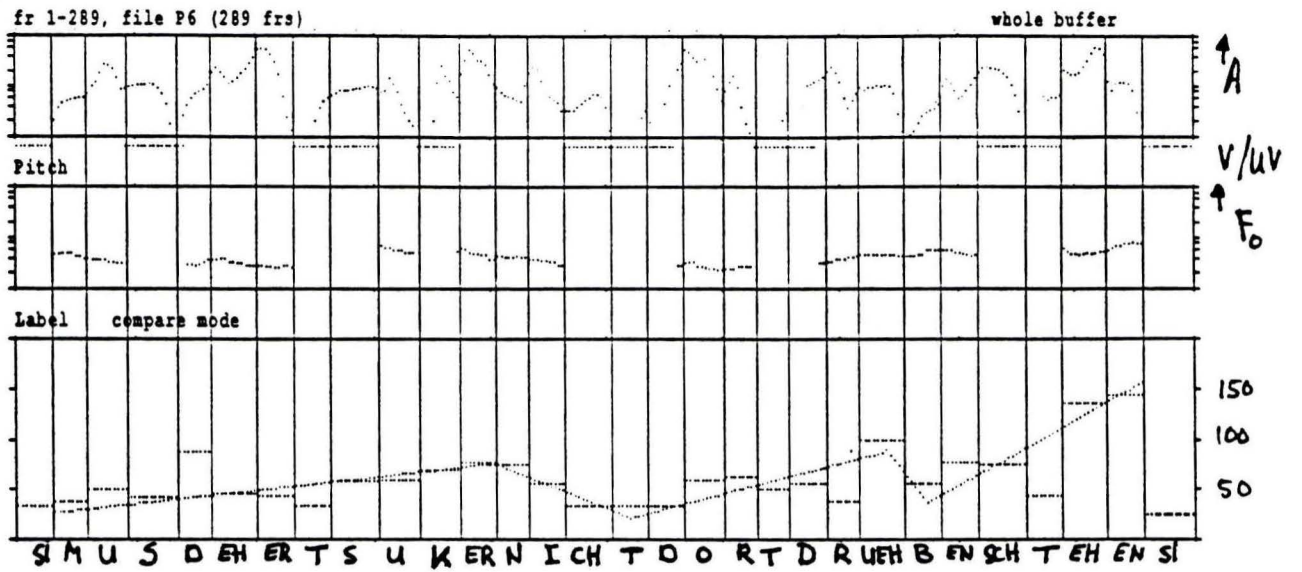
*Satzdauerwerte in Sekunden.*

	DS	O	P	A
1	2.42	1.765	1.762	1.712
2	1.89	1.452	1.467	1.574
3	2.51	1.711	1.699	1.790
4	1.74	1.200	1.190	1.100
5	1.59	1.020	1.046	1.114
6	2.89	1.730	1.702	1.774
7	1.81	1.171	1.223	1.194
8	1.95	1.482	1.483	1.468
9	2.08	1.299	1.305	1.333
10	2.69	1.683	1.702	1.690
11	1.82	1.070	1.083	1.088
12	1.81	1.120	1.089	1.135
13	1.96	1.351	1.347	1.401
14	2.26	1.581	1.578	1.747
15	2.63	1.551	1.567	1.545

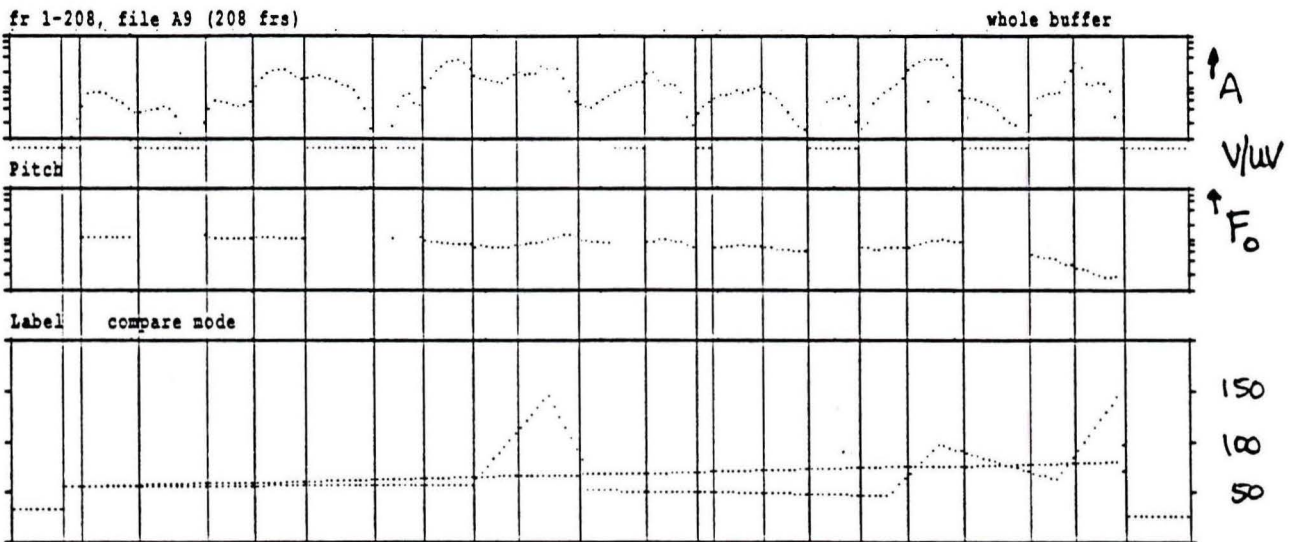
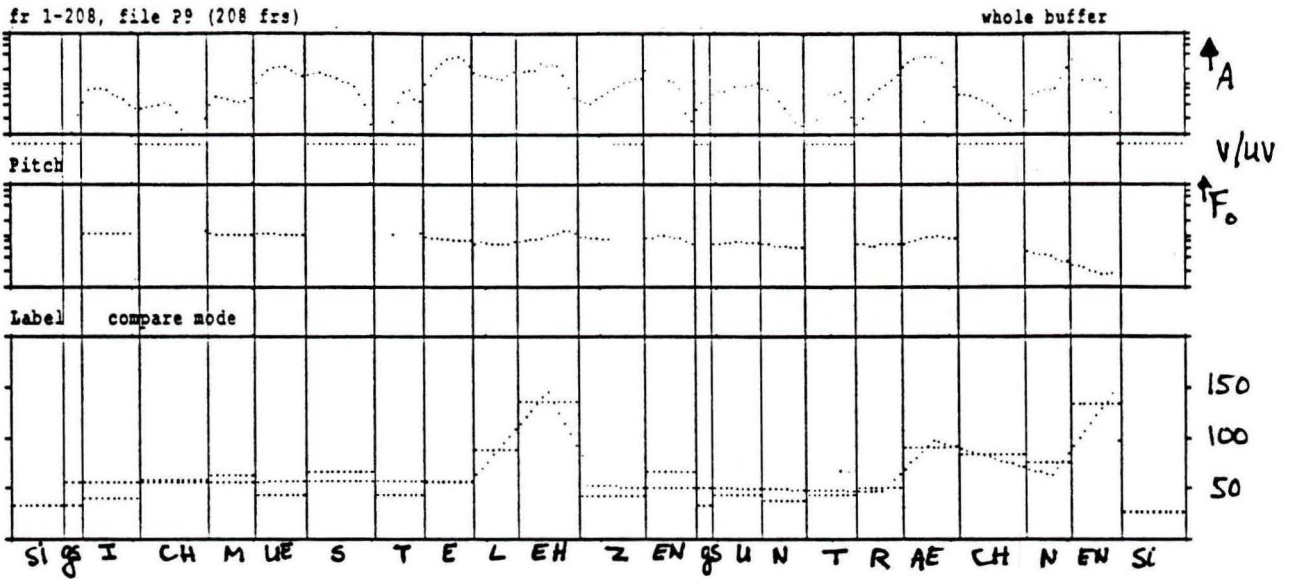
# 10 Anhang E: Beispiele der Stilisierungen und Ableiter



*Oben: Original Timingstruktur und Stilisierung, unten: Original Timingstruktur und Ableiter.*

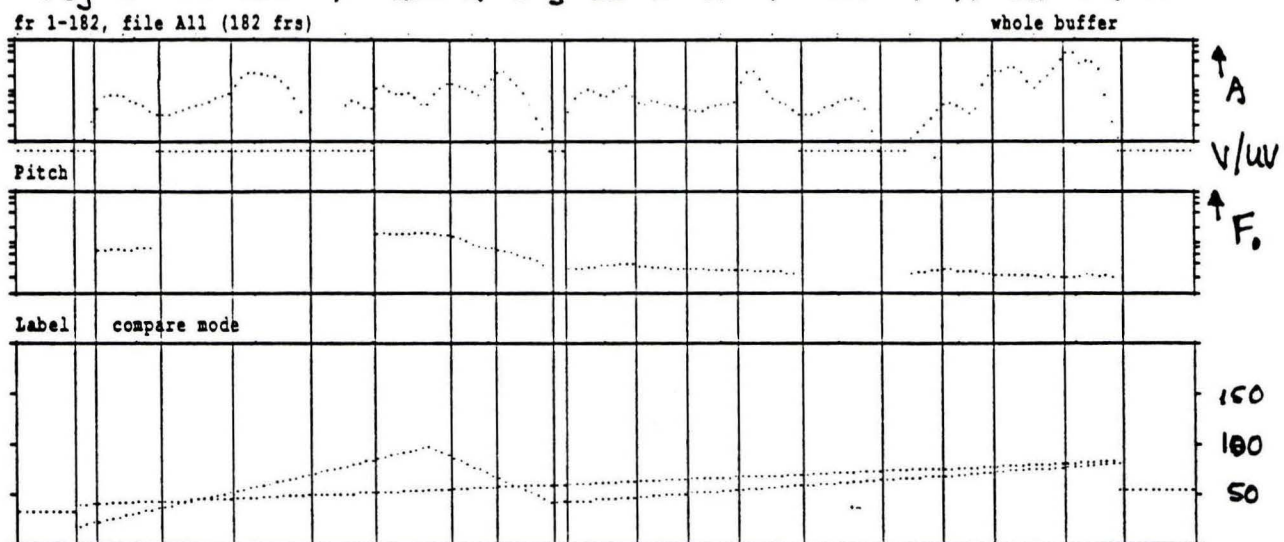
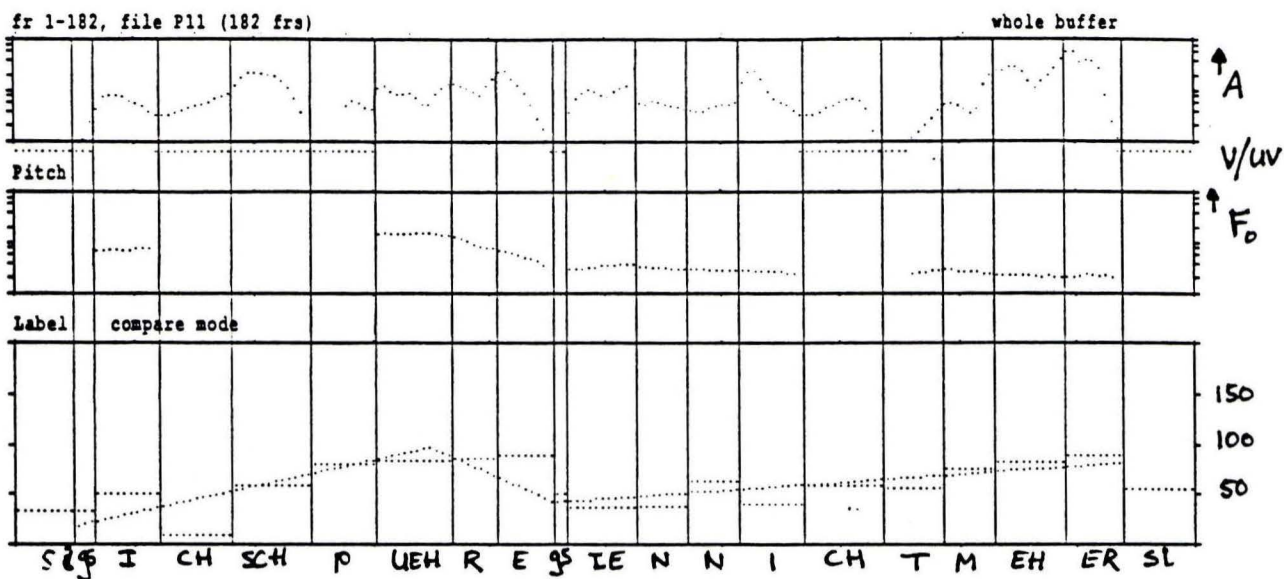


Oben: Original Timingstruktur und Stilisierung, unten: Stilisierung und Ableiter.



*Oben: Original Timingstruktur und Stilisierung, unten: Stilisierung und Ableiter.*





Oben: Original Timingstruktur und Stilisierung, unten: Stilisierung und Ableiter.