

SPRACHSEGMENTIERUNG MIT HILFE DER RHYTHMUSWAHRNEHMUNG

M. Köhlmann

Institut für Elektroakustik, Technische Universität München

1. Einleitung

Bezüglich der Wahrnehmung von fließender Sprache ist die Frage nach der zeitlichen Gliederung durch das Gehör noch weitgehend ungeklärt. Unter der Voraussetzung, daß Sprache einen Rhythmus besitzt, müßte es möglich sein, die vom Gehör vorgenommene Unterteilung in einzelne Segmente zu ermitteln. Es wurden deshalb psychoakustische Messungen des wahrgenommenen Rhythmus von Sprache durchgeführt. Eine Folge von rhythmischen Ereignissen beschreibt die zeitliche Anordnung der Segmente. Die Untersuchungen erfolgten an Sätzen in deutscher, englischer, französischer und japanischer Sprache und an rückwärts wiedergegebener Sprache. Es wurde versucht, die Meßergebnisse in einen Zusammenhang mit dem zeitlichen Lautheitsverlauf nach [4] zu bringen.

2. Versuchsdurchführung

In einer schallisolierten Meßzelle wurden den Versuchspersonen (Vpn) Sätze von einer Tonbandschleife diotisch über Kopfhörer (Beyer DT48 mit Entzerrer) dargeboten. Die Vpn hatten die Aufgabe, den wahrgenommenen Rhythmus der Sprache auf einer Morsetaste, welche an einen Prozeßrechner für die Auswertung der Antworten angeschlossen war, zu reproduzieren. Ein Schalldruckimpuls mit einer Dauer von 50  $\mu$ s und einem Spitzenschalldruckpegel von 90 dB diente als akustisches Kontrollsignal beim Nachtasten. Die Vp konnte die Darbietung eines Satzes und dessen Rhythmus-Reproduktion so oft wiederholen, bis sie mit dem Ergebnis zufrieden war. Die Vp mußte von jedem Satz 6 Rhythmus-Reproduktionen vornehmen. Die Sätze wurden in zufälliger Reihenfolge mit einer Lautheit von ca. 20 soneg (Spitzenwert) dargeboten. An den Messungen war eine Gruppe von 17 Vpn beteiligt, aus der jeweils 5 Vpn (bei den Sätzen in englischer und französischer Sprache je 3 Vpn) an einem Versuch teilnahmen. Die reproduzierten Rhythmen wurden, sofern sie in der Anzahl der Segmente übereinstimmten, zusammengefaßt und aus jeweils entsprechenden Zeitintervallen Zentralwerte und Wahrscheinliche Schwankungen berechnet. Im Folgenden werden zur Darstellung der reproduzierten Rhythmen die gemittelten Zeitintervalle, repräsentiert durch ihre Zentralwerte und Wahrschein-

lichen Schwankungen, zeitlich aneinandergereiht.

### 3. Sprachmaterial

Als Sprachmaterial in deutscher Sprache dienten 36 Sätze aus dem Marburger Satzverständnistest [3]. Davon wurden 6 Sätze zusätzlich für rückwärts wiedergegebene Sprache benutzt. In englischer und französischer Sprache wurden je 3, in japanischer Sprache 6 willkürlich gewählte Sätze verwendet.

### 4. Ergebnisse

Alle untersuchten deutschen Sätze wurden von den Vpn in eine jeweils einheitliche Anzahl von Segmenten unterteilt. Das heißt, daß interindividuelle Unterschiede sich nur in der Länge, nicht aber in der Anzahl der Segmente zeigten. Abb. 1 zeigt die derart gewonnene Segmentierung eines deutschen Satzes. Man sieht daran einen Zusammenhang zwischen dem Rhythmus des Satzes und dessen Silbenstruktur: Die Anzahl der Ereigniszeitpunkte stimmt mit der Anzahl der Silben überein. Alle untersuchten deutschen Sätze zeigten dieses Ergebnis. Im Gegensatz zu anderen Untersuchungen [1, 2], wo die Vpn jede Silbe markieren mußten, wurde hier spontan eine Zuordnung von rhythmischem Ereignis und Silbe getroffen. Die Auswahl der Vpn beeinflußt die rhythmische Struktur eines Satzes offenbar nur unwesentlich.

Abb. 2 zeigt die Segmentierung von 3 Sätzen, gewonnen durch 2 willkürlich gebildete Gruppen zu je 5 Vpn. Der Rhythmus unterscheidet sich in beiden Gruppen nur geringfügig. Zur Untersuchung des Zusammenhangs von Rhythmus und Lautheit wurden die gemessenen Rhythmen zeitlich auf das erste und letzte Maximum der entsprechenden Lautheitsverläufe normiert. In den meisten Fällen läßt sich eine Zuordnung von Ereigniszeitpunkt und Lautheitsmaximum treffen. Die tatsächliche Lage der Ereigniszeitpunkte im Lautheitsverlauf bzw. im Sprachsignal kann man aber auf diese Weise nicht ermitteln, da die Normierung weitgehend willkürlich gewählt ist und feste Bezugspunkte nicht vorhanden sind.

Beispiele für die Segmentierung von Fremdsprachen sind in Abb. 3 zu sehen. Ebenso wie im Deutschen ist auch hier eine rhythmische Struktur vorhanden, die eine zeitliche Gliederung ermöglicht. Alle Vpn verstanden Englisch, viele auch Französisch. Keine Vp besaß jedoch japanische Sprachkenntnisse. Da sich das Ergebnis mit der japanischen Sprachprobe nicht merklich von den übrigen unter-

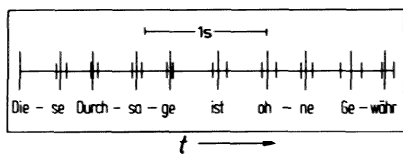


Abb. 1. Beispiel für Sprachsegmentierung. Zentralwerte und Wahrscheinliche Schwankungen aufgrund der reproduzierten Rhythmen.

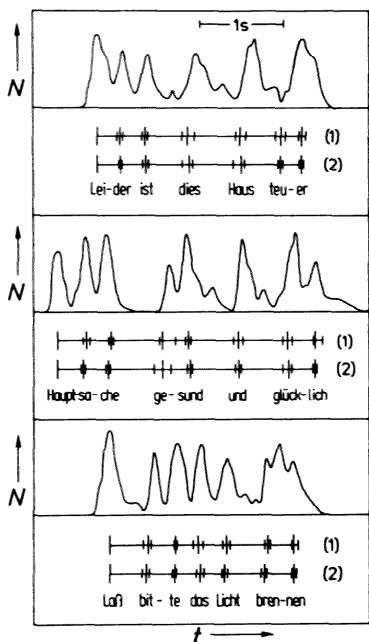


Abb. 2. Segmentierung aus reproduziertem Rhythmus und Lautheitsverlauf von 3 deutschen Sätzen [(1), (2): Gruppen mit je 5 Vpn].

Abb. 4. Segmentierung und Lautheitsverlauf eines rückwärts (R) und vorwärts (V) wiedergegebenen Satzes ((R) zeitlich invertiert dargestellt.)

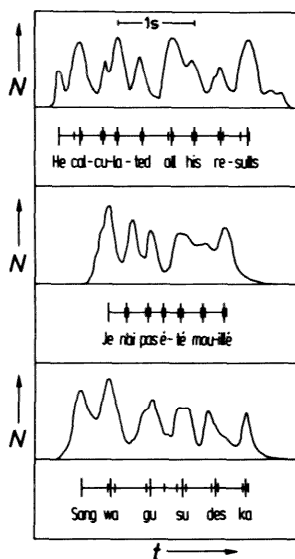
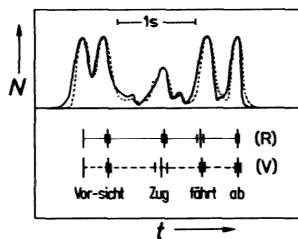


Abb. 3. Segmentierung und Lautheitsverlauf von je einem Satz in englischer, französischer und japanischer Sprache.



scheidet, wird geschlossen, daß die Segmentierung primär nicht an die Sprachverständlichkeit gebunden ist. Die gezeigten Beispiele weisen ebenso wie im Deutschen einen Zusammenhang zwischen Rhythmus und Lautheit auf; jedoch ist auch hier die Beschreibung der Rhythmuswahrnehmung allein durch Lautheitsschwankungen unvollständig.

Ein Beispiel für rückwärts wiedergegebene, also ebenfalls unverständliche Sprache zeigt Abb. 4. Segmentierung und Lautheitsverlauf des rückwärts wiedergegebenen Satzes sind darin im Vergleich mit demselben, vorwärts wiedergegebenen Satz zeitlich invertiert dargestellt. Es zeigt sich, daß die Sprache sowohl vorwärts als auch rückwärts wiedergegeben in die gleiche Anzahl von Segmenten, die sich in ihrer Länge nur wenig unterscheiden, gegliedert wird. In beiden Fällen läßt sich eine Zuordnung von Ereigniszeitpunkten und Lautheitsmaxima finden. Eine rückwärts wiedergegebene Sprachäußerung besitzt somit näherungsweise den zeitlich invertierten Rhythmus derselben, vorwärts wiedergegebenen Sprachäußerung.

### 5. Zusammenfassung

Sprache läßt sich mit Hilfe der Rhythmuswahrnehmung in zeitliche Abschnitte mit einer Dauer bis zu einigen 100 ms unterteilen. Mindestens im Deutschen spiegelt sich darin die Silbenstruktur wider. Die Sprachverständlichkeit spielt dabei keine oder nur eine ganz untergeordnete Rolle. Die Lautheit trägt offenbar maßgeblich zur Entstehung der rhythmischen Ereignisse bei, reicht aber zur vollständigen Beschreibung der Meßergebnisse nicht aus.

### Literatur

- [1] Allen, G.D. (1972) The location of rhythmic stress beats in English. Part I/II. *Language & Speech* 15, 72-100, 179-195.
- [2] Folkins, J.W., Miller, C.J., Minifie, F.D. (1975) Rhythm and syllable timing in phrase level stress patterning. *Journ. of Speech and Hearing Research* 18, 739-753.
- [3] Niemeyer, W., Beckmann, G. (1962) Ein sprachaudiometrischer Satztest. *Arch. Ohr.-Nas.-u. Kehlk.-Heilk.* 180, 742.
- [4] Zwicker, E. (1982) *Psychoakustik*. Springer Verlag, Heidelberg.

*Frau Prof. Y. Yamada (Tokio) sei für die japanischen Sprachaufnahmen und Herrn Prof. Dr.-Ing. E. Terhardt für die konstruktive Kritik beim Abfassen des Manuskripts gedankt. Diese Arbeit wurde von der Deutschen Forschungsgemeinschaft im Rahmen des Sonderforschungsbereich 50 "Kybernetik" gefördert.*