

## Physiologisch-artikulatorische Korrelate zur Realisierung von Akzentmustern und Intonationstypen

C. Oppen-Rhein, B.J. Kröger

Institut für Phonetik, Universität zu Köln

### Einleitung

Die Einbettung prosodischer Phänomene ist für die Qualität einer artikulatorisch-phonatorisch basierten Sprachsynthese unerlässlich. Ein Ansatz zur Modellierung von einfachen terminalen und interrogativen Sätzen mit unterschiedlicher Akzentstruktur (Satz-, Kontrastakzent) für ein artikulatorisches Sprachsynthesystem [3] wird vorgestellt. Als akustische Komplexphänomene manifestieren sich die prosodischen Phänomene in den physiologisch-phonatorischen Parametern Stimmlippenspannung und pulmonaler Druck. Regeln für diese Parameter werden mittels Analyse durch Resynthese gewonnen und führen zur Generierung eines Intonationsmodells des Deutschen.

### Datenbasis

Die für diese Arbeit verwendeten Testsätze mit unterschiedlichem Intonationstyp und Akzentmuster ("Er malt", "Er niest", "Lars duldet") ergaben sich aus der Minimalpaarmethode. In den vorgestellten Minimalpaaren ist bei gleicher segmentaler Struktur die prosodische Form bedeutungsunterscheidend. Als akustische Datenbasis für die Resynthese dienten Äußerungen, die von einem männlichen Sprecher des Standarddeutschen aufgenommen wurden.

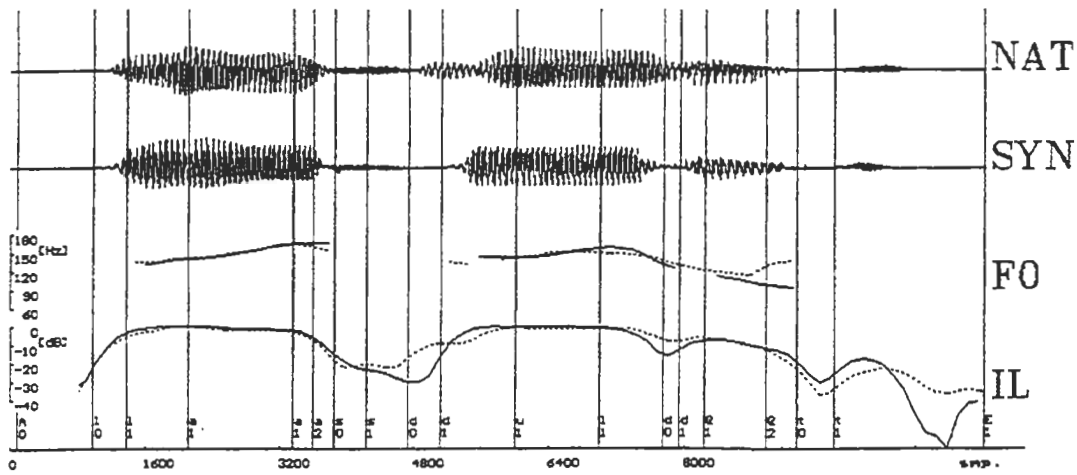
### Artikulatorisch-phonatorisches Sprachsynthesystem

Auf der Basis artikulatorischer und phonatorischer Steuerparameter liefert das Sprachsynthesystem eine möglichst natürliche Modellierung des menschlichen Sprachproduktionsmechanismus. Zur Realisierung prosodischer Phänomene dienen als Input nicht die akustischen Größen (F<sub>0</sub>, Intensitätslevel), sondern die physiologisch-artikulatorisch basierten Steuerparameter Stimmlippenspannung, glottaler Öffnungsgrad und pulmonaler Druck. Stimmlippenspannung und glottaler Öffnungsgrad werden durch ein selbstschwingendes Zwei-Massen-Modell der Glottis [2] simuliert. Die Grundlage des Intonationsregelmodells bildet das segmentale Regelmodell, welches die räumliche und zeitliche Organisation von artikulatorischen und phonatorischen Abläufen gewährleistet [4].

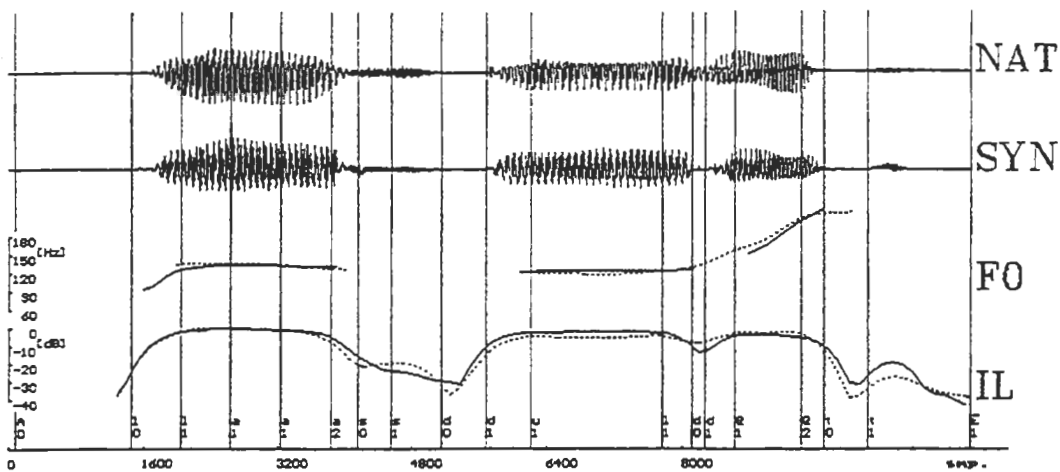
### Resynthese

Innerhalb der Resynthese werden die prosodischen Merkmale Grundfrequenz, Intensität und Dauer in Betracht gezogen. Die Relation der Modellparameter ist wie im Ishizaka/Flanagan-Modell [2] die folgende: "q" entspricht der Stimmlippenspannung und ist dimensionslos, "p<sub>s</sub>" entspricht dem Lungendruck in cmH<sub>2</sub>O. Das Synthesemodell ermöglicht es, ein natürliches und ein zugehöriges Resynthesesignal einander gegenüberzustellen (Abb.1, exemplarisch: "Lars duldet").

Veränderungen der Stimmlippenspannung beschränken sich auf sonorante Laute, d.h. Vokale, Diphthonge, Halbvokale, Liquide und Nasale. Veränderungen des pulmonalen Drucks können bei Sonoranten und Obstruenten auftreten, solange sie einen bestimmten durch die Regelkomponente festgelegten Grenzwert nicht unterschreiten, da in diesem Fall kein Friktions- oder Aspirationsgeräusch entstehen würde. Der Prozeß der Resynthese ist interaktiv.

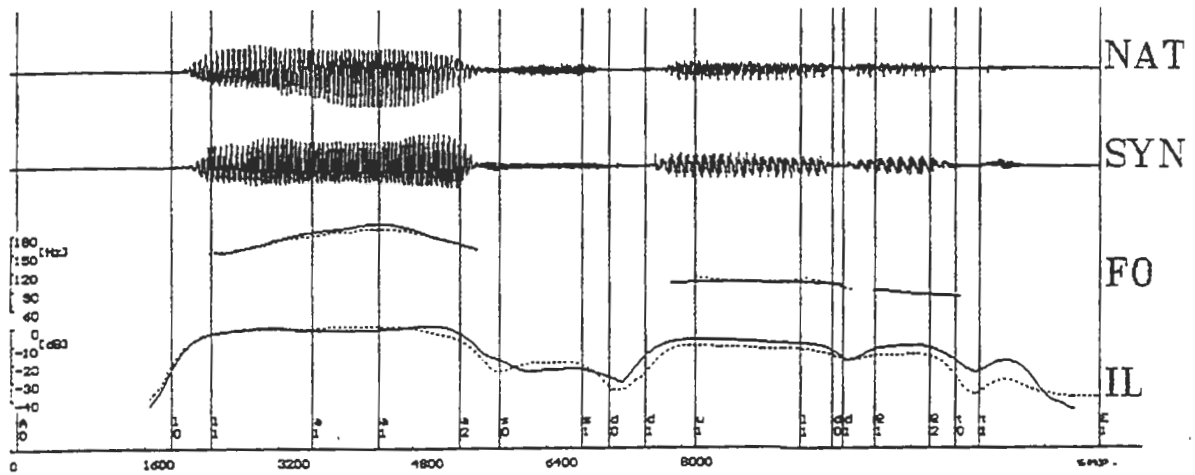


(a)

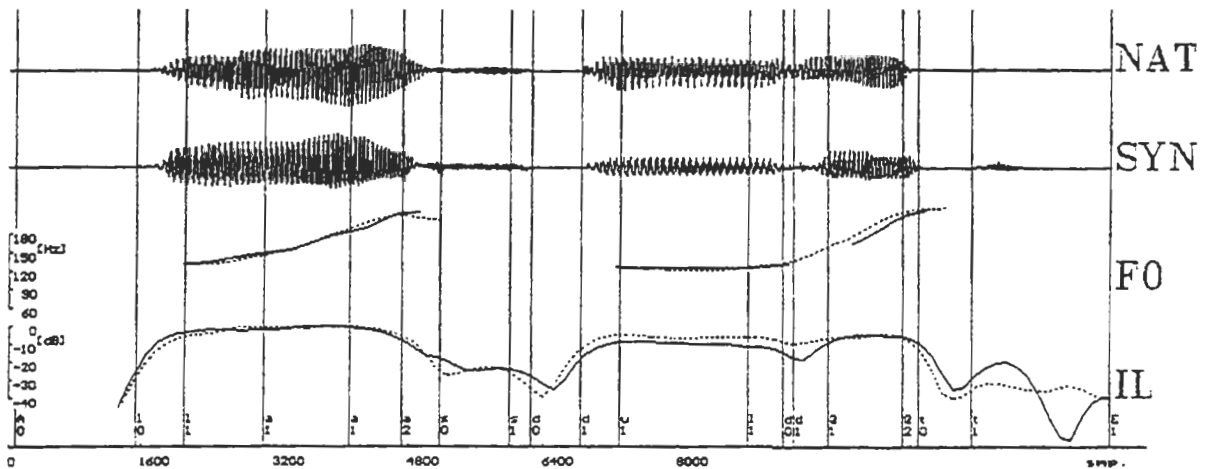


(b)

Abbildung 1 Oszillogrammdarstellung der natürlichen (NAT) und synthetischen Signale (SYN) mit zugehörigen Grundfrequenz (FO)- [Hz] und Intensitäts (IL)- [dB] verläufen (natürlich: gestrichelt, synthetisch: durchgezogen) der verschiedenen Testsätze "Lars duldet": (a) Satzakzent, terminal, (b) Satzakzent, interrogativ, (c) Kontrastakzent, terminal, (d) Kontrastakzent, interrogativ. Die senkrechten Striche zeigen die Label. Anfang und Ende der Äusserung werden durch das Anfangslabel (A) und das Endlabel (E<sub>1</sub>) markiert, welche durch ihre Targeteinstellung die prä- und postphonorische Stellung (Atemstellung) vorgeben, dazwischen erscheinen die Label zur Kennzeichnung der Eingabesymbole. Auf der horizontalen Achse wird der zeitliche Verlauf in Sampeln angegeben ( $f_s=20\text{kHz}=1\text{sec.}$ ).



(c)



(d)

### Intonationsmodell

Das Ziel des Intonationsmodells ist die Modellierung möglichst natürlicher Intonationskonturen für abgeschlossene zwei- bzw. dreisilbige Äußerungen mit einer minimalen Anzahl von Transkriptionssymbolen auf der Eingabeebene. Es gewährleistet die Realisierung des Intonationstyps (terminal/interrogativ) und des Hauptakzents (Satz-/Kontrastakzent). Dies setzt eine zusätzliche prosodische Komponente zum bisher bestehenden Regelsystem voraus. Die Targetwerte für Stimmlippenspannung und pulmonalen Druck ergeben sich aus den in der Resynthese erhobenen Mittelwerten. Sie sind durch die räumliche und zeitliche Bestimmung auf einen einzelnen Punkt in der sequentiellen Abfolge der Laute beschränkt. Durch Anwendung dynamischer Regeln werden die Trajektorien zwischen diesen einzelnen Zielpunkten erzeugt und damit auch die Intonationskonturen im akustischen Modell.

## Resultate

Als wichtigste aus der Resynthese ermittelte Ergebnisse, die als Gesetzmäßigkeiten in die Intonationsregeln einfließen, können folgende festgehalten werden:

1. Die Grundfrequenz wird vorwiegend durch die Stimmlippenspannung beeinflusst. Mit steigender Stimmlippenspannung nimmt die Grundfrequenz zu.
2. Intensität variiert in erster Linie mit dem pulmonalen Druck. Sie ist somit primär eine Funktion des pulmonalen Drucks.
3. Darüberhinaus beeinflussen sich Stimmlippenspannung und pulmonaler Druck gegenseitig. Diese Wechselwirkung hat wiederum Auswirkungen auf Grundfrequenz und Intensität. So kann der pulmonale Druck, wenn auch nur geringfügig, Veränderungen der Grundfrequenz hervorrufen. Insbesondere wird bei einer Steigerung der Stimmlippenspannung ein stärkerer pulmonaler Druck erforderlich, um eine konstante Intensität zu bewirken. Eine extrem starke Stimmlippenspannung, wie sie z.B. in interrogativen Intonationsverläufen oder bei der Realisierung des Kontrastakzents zu beobachten ist, erniedrigt die Intensität. Dennoch ist in natürlich gesprochenen Sätzen kein Abfall der Intensität zu erkennen. Um die erforderliche Intensität in den Synthesesignalen beizubehalten, ist eine Erhöhung des pulmonalen Drucks erforderlich. Aufgrund dieser Gesetzmäßigkeiten erfolgt bei der Resynthese zunächst eine Angleichung der Stimmlippenspannung und erst dann eine Angleichung des pulmonalen Drucks.

## Diskussion

Eine Verifizierung des Intonationsmodells wurde mittels eines Perzeptionsexperiments durchgeführt. Das Sprachproduktionsmodell ist in der Lage, aufgrund des erstellten Intonationsmodells vorgegebene Intonationstypen und Akzentmuster über die phonatorischen Steuerparameter zu generieren. Eine Modifikation der Regeln muß jedoch bezüglich des Satzakzents stattfinden, indem die Steuerparameter Stimmlippenspannung und Dauer als wesentliche Faktoren die Hervorhebung der Satzakkentsilbe bestimmen. Der Aspekt der Dauer erfordert zusätzliche Untersuchungen, vor allem die Kennzeichnung satzfinaler Längung, Reduktionsphänomene in nicht-akzentuierten Silben sowie die Längung in akzentuierten Silben.

## Literatur

- [1] Bannert, R. (1985). Fokus, Kontrast und Phrasenintonation im Deutschen. *Zeitschrift für Dialektologie und Linguistik* 52, 289-305.
- [2] Ishizaka, K., Flanagan, J.L. (1972). Synthesis of voiced sounds from a two-mass-model of the vocal cords. *Bell System Technical Journal* 51, 1233-1268.
- [3] Kröger, B.J. (1992). Minimal rules for articulatory speech synthesis. In: *Signal Processing VI: Theories and applications* (ed. J. Vandewalle, R. Boite et al.), Amsterdam: Elsevier, 331-334.
- [4] Kröger, B.J., Opgen-Rhein, C. (1992). Ein phonembasiertes Produktionsmodell für artikulatorische Sprachsynthese. *DAGA'92*, 533-535.
- [5] Monsen, R.B., Engebretson, A.M., Vemula, R.N. (1978). Indirect assessment of the contribution of subglottal air pressure and vocal-fold tension to changes of fundamental frequency in English. *Journal of the Acoustical Society of America* 64, 65-80.
- [6] Stock, E., Zacharias, C. (1982). *Deutsche Satzintonation*. Leipzig: VEB Verlag Enzyklopädie.