

# VOKALE UND IHRE PHYSIKALISCHEN MERKMALE. I. LAUTIDENTITÄT UND FORMANTMUSTER.

Dieter Maurer

Universitätsspital Zürich-CH. Neurologische Klinik

Andreas Klinkert

Universität Freiburg-CH. Institut für Informatik

Forschung unterstützt von: Olga Mayenfisch Stiftung. Theodor und Ida Herzog - Egli Stiftung. Jubiläumstiftung Schweizerische Lebensversicherungs- und Rentenanstalt

## Einleitung

In der akustischen Phonetik wird üblicherweise davon ausgegangen, dass sich die Lautidentität von Vokalen physikalisch in charakteristischen Mustern von Resonanzen, sogenannten Formantmustern, ausdrückt. Es ist allerdings in der Literatur ausführlich beschrieben, dass die Muster eines Lautes zum Teil stark variieren. Diese Variationen werden meist als Folge verschiedener Sprechergruppen (Männer, Frauen, Kinder), als Folge verschiedener Vokalisationsarten, oder als Folge von Koartikulation interpretiert. Studien an synthetischen Vokalen weisen zudem darauf hin, dass die Formantmuster eines Vokals auch unabhängig von Sprechergruppe, Vokalisationsart und Lautkontext variieren können: Die tieferen Formanten erweisen sich in der Synthese als abhängig von  $F_0$ , und die für eine Lautidentität relevante Zahl der Formanten kann verschieden sein. – Wir untersuchten Formantmuster von isolierten natürlichen und synthetischen Vokalen im Sinne von Resonanzcharakteristika der entsprechenden Signale, und fanden sowohl die Abhängigkeit der tieferen Formanten von  $F_0$  wie die Variation der Formantzahl bestätigt. Darüber hinaus erwies sich der Versuch, Formantmuster einer Vokalidentität zuzuordnen, aber grundsätzlich als problematisch: das Verhältnis von  $F_0$  und tieferen Formanten erwies sich als unsystematisch (weder linear noch monoton); die Variation der Formantzahl liess sich nicht auf eine spektrale Integration zurückführen; sehr ähnliche Formantmuster fanden sich bei verschiedenen Vokalen (Ambiguität der Formantmuster); Vokale wurden auch dann erkannt, wenn die entsprechenden Spektren keine deutliche Resonanzstruktur zeigten; Formanten waren oft nicht bestimmbar. – Einige dieser Befunde wurden schon in früheren Studien vorgestellt (Einzelstudien vgl. [3-5]; ausführliche Dokumentation anhand einer älteren Vokalsammlung vgl. [6]; exemplarische Darstellung mit Illustrationen vgl. [7]). In diesem Beitrag werden die Ergebnisse einer systematischen Prüfung der Entsprechung von Vokalen und Formantmustern anhand einer neuen umfangreichen Vokalsammlung dargestellt (vgl. auch [8]). Parallel dazu werden im Beitrag "Vokale und ihre physikalischen Merkmale II" unsere jüngsten Versuche, zu einer robusten Aussage über das Verhältnis von Spektrum und Lautidentität zu kommen, zur Diskussion gebracht.

## Fragestellung und Methode

Fragestellung: Welche Entsprechung von Vokal und Resonanzmustern eines Signals zeigt sich, wenn gilt: i) die Vokale werden stimmhaft und isoliert gebildet; ii)  $F_0$  und Lautstärke der Vokalisationen werden variiert; iii) Ein- und Ausschwingvorgänge werden gelöscht; iv) die Signale zeigen ein hohes Ausmass an spektraler Konstanz (Ausschluss von Transitionen); v) die wahrnehmungsmässige Zuordnung von Signalen und Vokallauten in einem Identifikationstest ist sehr ausgeprägt.

Untersuchte Vokalsammlung: Schweizerdeutsche Vokale /i,y,e/,E,Ø,o,u/ (Zürcher Dialekt); Anzahl untersuchter Klänge = 18'674 Aufnahmen natürlicher Vokale, stimmhaft und isoliert gebildet (Vokalisationsdauer = 1.5sec), und 250 synthetische Vokale; Variation von  $F_0$  und Lautstärken (Bereich  $F_0$  = 70 - 850 Hz); Anzahl Sprechende = 99 (35 Männer, 44 Frauen, 20 Kinder).

Bearbeitung der Signale: Extraktion von mittleren Lautsegmenten (0.7sec); Fourier und LPC Analyse; interaktive Formantbestimmung; Kontrolle der Konstanz von Spektrum und spektraler Hüllkurve über die ganze Dauer eines Lautsegmentes; Prüfung der Identifikation der Lautsegmente (Identifikationstests).

Zum Verständnis der Untersuchung als ganzes ist zu beachten, dass unter "Formanten" immer nur Resonanzen eines Signals (nicht aber Resonanzen des Vokaltraktes) verstanden und entsprechend untersucht wurden. Die folgenden Ausführungen beziehen sich derart auf die akustische Beschreibung von Vokalen.

## Ergebnisse

**1. Eine systematische und statistische Untersuchung der Entsprechung von Formantmustern und Vokalen ist aus methodischen Gründen nicht möglich.** Es gibt bis heute keine vollständig objektivierte Methode der Bestimmung von Formanten aus einem Signal. Formanten werden deshalb in der Regel interaktiv ermittelt, wobei Kenntnisse der akustischen Phonetik seitens der Untersuchenden eine wichtige Rolle spielen. Die Problematik der Formantmessung nimmt dabei mit steigenden Werten für  $F_0$  zu. Als grobe Regel kann gelten, dass die Werte für tiefere Formanten ab  $F_0 > 200$ Hz, diejenigen für höhere Formanten ab  $F_0 > 400$ Hz unzuverlässig werden. Eine Formantbestimmung für  $F_0 > 400$ Hz führt deshalb bei vielen Vokalisationen zu keinen aussagekräftigen Werten. Die Erkennbarkeit von Vokalen übersteigt aber bei weitem diese Grenze: isolierte Vokale können bis auf  $F_0 \approx 700$ Hz deutlich erkannt werden, und eine Unterscheidung der Vokale /u-Ø-i/ findet sich bis  $F_0 \approx 850$ Hz.

**2. Für natürliche Vokalisationen, deren Formantmuster im Sinne deutlicher spektraler Spitzen angezeigt sind, gibt es keine Konstanz der für die Lautidentität massgebenden Formantzahl.** Insbesondere lassen sich hintere "Einformanten-Vokale" (nur eine deutliche spektrale Spitze < 2kHz) und vordere "Zweiformalanten-Vokale" (nur zwei deutliche spektrale Spitzen < 4kHz) beobachten. Eine allgemeine spektrale Integration zweier nahe beieinanderliegender Formanten zu einem dazwischen liegenden konnte experimentell nicht bestätigt werden (vgl. auch [9,10]).

**3. Für natürliche Vokalisationen, deren Formantmuster im Sinne deutlicher spektraler Spitzen angezeigt sind, verhalten sich die tieferen Formanten abhängig von  $F_0$ .** Das Ausmass

der Verschiebung der tieferen Formanten (in Hz) kann dabei dasjenige von F0 weit übertreffen. Umgekehrt führt das Beibehalten von Formanten bei einer Änderung von F0 oft zu einer deutlichen Lautverschiebung.

**4. Die Abhängigkeit der tieferen Formanten von F0 ist non-linear.** Das Ausmass der Formantverschiebungen variiert in Abhängigkeit der Frequenz F0. Schwache und uneinheitliche Verschiebungen finden sich für F0<200Hz, äusserst ausgeprägte hingegen für F0=200-400Hz. Darüber ist eine Aussage auf Grund der allgemeinen Bestimmungsproblematik der Formanten schwer zu treffen.

**5. Die Abhängigkeit der tieferen Formanten von F0 ist non-monoton.** Das Ausmass der Formantverschiebungen variiert in Abhängigkeit der Vokallaute. Schwache und uneinheitliche Verschiebungen finden sich für /E/ und /Ø/. äusserst ausgeprägte hingegen für /e/, /l/ und /o/. Eine meist parallele Verschiebung der Formanten mit F0 zeigen /i/, /y/ und /u/.

**6. Die Abhängigkeit der tieferen Formanten von F0 widerspricht dem angenommenen generellen Unterschied der Formantmuster verschiedener Sprechergruppen: Es gibt keine allgemeine Beziehung von Vokaltraktgrösse und tieferen Formanten.** Wenn Männer, Frauen und Kinder auf derselben Grundtonhöhe vokalisieren, so findet sich kein grundsätzlicher Unterschied der tieferen Formanten.

**7. Für natürliche Vokalisationen, deren Formantmuster im Sinne deutlicher spektraler Spitzen angezeigt sind, verhalten sich die Formantmuster mehrdeutig - sie entsprechen verschiedenen Vokallauten (Ambiguität der Formantmuster).** Drei Beobachtungen sind dabei von besonderer Bedeutung: Die Mehrdeutigkeit der Formantmuster betrifft alle für die Lautidentität massgebenden Formanten: ein bestimmtes Formantmuster kann nicht nur zwei, sondern mehreren verschiedenen Vokalen entsprechen (multiple Ambiguität); quasi-identische Formantmuster finden sich für Vokale mit kleiner wie mit grosser phonetischer Distanz.

**8. Alle Beobachtungen sind in der Synthese nachbildbar.** Sie können deshalb nicht als Folge der Problematik der Formantmessung interpretiert und relativiert werden.

**9.) Synthetische Vokale können erkannt werden, auch wenn sie "anomale" Spektren - Spektren mit Eigenschaften weit ab von Resonanzmustern, wie sie in Formanttabellen dargestellt werden - aufweisen.** Insbesondere können Vokale deutlich erkannt werden, auch wenn sie über grössere Frequenzbereiche keine eigentliche Resonanzstruktur besitzen.

## Folgerungen

- Die These der direkten Entsprechung von Formantmustern und Vokalidentität ist durch den Nachweis der Ambiguität der Muster für isolierte Vokale falsifiziert. Formanten sind nicht das physikalische Merkmal des Vokallauts.

- Die These der Entsprechung von Formantverhältnissen und Vokallauten (Konzept der Normalisation) kann auf Grund der methodischen Schwierigkeiten der Formantbestimmung zur Zeit experimentell nicht für alle Fälle deutlich wahrnehmbarer natürlicher Laute geprüft werden.

- Die methodische Problematik, zusammen mit den oben genannten Anzeichen eines unsystematischen Verhältnisses von F0 und Formantfrequenzen und einer Variation der Formantzahl, lassen grundsätzliche Zweifel am Versuch aufkommen, Vokallaute akustisch über Formantmuster oder Formantverhältnisse zu beschreiben.

- Die Problematik der Entsprechung von Formanten und Vokalen zeigt sich experimentell äusserst deutlich für quasi-stati-

sche Klänge. Dynamische Konzepte stellen deshalb keine Alternative dar.

- Die in diesem Beitrag aufgeführten Beobachtungen stellen Phonation und Artikulation in ein neues Licht: Die Abhängigkeit der Formanten von F0 widerspricht der Unabhängigkeit der beiden Vorgänge, und die "Ambiguität" der Formanten widerspricht der Zuordnung von spezifischen Artikulationsstellungen zu einem Vokallaut. (Zum Verhältnis von Artikulation, Phonation und Vokalspektrum, vgl. auch [11,12].)

- Die Beobachtungen bekräftigen frühere Argumentationen, dass Formantmuster für die Perception von Vokalen keine direkte Bedeutung besitzen [13].

- Die Zuordnung von wahrgenommenem Vokallaut und physikalischen Merkmalen verlangt deshalb nach einem neuen robusten Ansatz (vgl. Parallelbeitrag « Vokale und ihre physikalischen Merkmale II »).

## Literatur

- [1] Traumnüller, H. (1985): The role of the fundamental and the higher formants in the perception of speaker size, vocal effort, and vowel openness. In PERILUS No. 4 (Inst.Linguist.Univ.Stockholm), pp. 92-102.
- [2] Delattre, P., Liberman, A.M., Cooper, F.S., and Gerstman, L.J. (1952): An experimental study of the acoustic determinants of vowel color; observations on one- and two-formant vowels synthesized from spectrographic patterns. Word 8, 195-210.
- [3] Maurer, D., Cook, N., Landis, T. & d'Heureuse C. (1992): Are measured differences between the formants of men, women and children due to F0 differences? J. Int. Phonet. Assoc., 21, 66-79.
- [4] Maurer, D. & Landis, T. (1995): F0-dependence, number alteration, and non-systematic behaviour of the formants in German vowels. Int. J. Neurosci., 83, 25-44.
- [5] Maurer, D. & Landis, T. (1996): Intelligibility and spectral differences in high pitched vowels. Folia phoniat. logopaed., 48, 1-10.
- [6] Maurer, D. (1994): Über den Vokal. Band I: Kritik der akustischen Theorie der stimmhaften Sprachlaute. Band II: Materialien. Konstanz, Hartung-Gorre Verlag.
- [7] Maurer, D. (1997): Arguments against formants - the descriptive problem of acoustic phonetics. Proceedings of the Journées d'Etudes Linguistiques 1997, University of Nantes (FR), 106-111.
- [8] Maurer, D., and Klinkert, A. (1997). Fourier spectra and formant patterns of German vowels produced at F0 of 70-85 Hz. J. Acoust. Soc. Am. 101, 3112.
- [9] Assmann, P.F. (1991): The perception of back vowels: Center of gravity hypothesis. Q. J. Exp. Psychol. 43A, 423-448.
- [10] Fahy, R.P., Diehl, R.L., and Traumnüller, H. (1996): Perception of back vowels: Effects of varying F1-F0 Bark distance. J. Acoust. Soc. Am. 99, 2350-2357.
- [11] Maurer, D., Gröne, B., Landis, T., Hoch, G. & Schönle, P.W. (1993): Reexamination of the relation between the vocal tract and the vowel sound with EMA in vocalizations. Clinical Linguistics & Phonetics, 7, 129-143.
- [12] Maurer, D., Hess, M. & Gross, M. (1996): High-speed imaged vocal fold vibrations and larynx movements within vocalizations of different vowels. Ann. Otol. Rhinol. Laryngol., 105, 975-981.
- [13] Bladon, A. (1982): Arguments against formants in the auditory representation of speech. In Carlson, R. & Granström, B. (Eds.): The Representation of Speech in the Peripheral Auditory System. New York. Elsevier, 95-102.

## Adresse für Korrespondenz:

Dieter Maurer, Dr. phil. I  
USZ, Neurologische Klinik, Neuropsychologische Abteilung  
Haldenbach D1 - CH-8091 Zürich  
Tel.: 41-1-255 55 71 Fax: 41-1-255 44 29  
Email: dmaurer@npsy.unizh.ch  
Internet: <http://www.unizh.ch/neuro/Maurer/index.htm>