

*Balázs Huszka (Veszprém):*

# Heuristischer Zugriff auf die Dynamik des Sprechschalls: Zur akustischen Typologie des Schwas im Deutschen in den Sequenzen $v_{[\theta]}c_{[s]}$ und $v_{[\theta]}c_{[ʃ]}$

## 1 Einleitung

### 1.1 Problemhintergrund

Wenn man die deutschbezogene Fachliteratur der Realisierung kortikaler Sprachlautrepräsentationen (Neureme) als Sprechlaute, d.h. als in akustische/nullakustische<sup>1</sup> Korrelate (Akuseme) der sich quasi-stereotypisch (in approximativ gleicher Form) zusammensetzenden, vollziehenden und wiederholenden kinetischen Komplexe (Kimeme) der Sprechorgane (vgl. Bolla 1981: 17f.) kodierete sprachliche Informationsstruktur übersieht (Wängler 1981, Bolla/Valaczkai 1986, Valaczkai 1998, um lediglich Monographien zu erwähnen), konfrontiert er mit den nachstehenden Grundproblemen:

1. Nebst der detaillierten Darstellung spezifischer Kalibrationen und Konstellationen der Sprechorgane v.a. in Röntgenbildern, Palato-, Linguo- und Labiogrammen sind die komplexen akustischen Vorgänge weniger resp.
2. auf Grund lediglich dreier Parameter, der relativen/spezifischen Bildungsdauer, der Frequenz und der Intensität definiert worden. Solche für Ausformung des definitiven Klang-/Geräuschbildes signifikanten Komponenten wie Elongation, Energie oder Harmonizität sind deutlich unterrepräsentiert.
3. Der über phonotaktische (evtl. etwaige perzeptuelle, vgl. Whalen 1991, Gósy 1989, nach Gósy 1995: 21) Relevanz verfügenden akustischen/nullakustischen Projektion koartikulatorischer Phänomene (Lautübergänge, Assimilationen) und
4. der Position (In-, An-, Auslaut, (vor und nach) betonte(r)/unbetonte(r) Silbe etc.) bzw. dem Kontext im Lautkörper fällt (im Verhältnis zu ihrer Bedeutsamkeit) ebenfalls wenig Aufmerksamkeit zu.

---

<sup>1</sup> Vértés O. (1982: 155) stellt von einer Erkenntnis Fants (1960: 218) ausgehend fest, dass die Akustik immer ihre artikulatorische Grundlage habe, nicht aber umgekehrt: Die Artikulation solle sich nicht unbedingt akustisch manifestieren. Das Phänomen lässt sich e.g. bei Explosivlauten unschwer beobachten, wo sich an die labiale Tätigkeit der Verschlussbildung (resp. in der stummen Haltephase des Verschlusses) die offenbar nicht außerhalb der Artikulation erfolgt, *in praxi* kein (hör- und registrierbares) Schallereignis anknüpft.

5. Die bisher veröffentlichten numerischen/prozentualen Werte der akustischen Struktur der Sprechlaute sind nicht oder zumindest nicht ausreichend dynamisch, wo doch die interne energetische Fluktuation dies erforderte.
6. Die zur Schau gestellten dynamischen akustischen Registrate (Oszillogramm, Spektrogramm etc.) können über alle wesentlichen strukturellen Veränderungen in der Zeitfolge keine Rechenschaft ablegen: Laut der phonetischen Interpretation der (aus der Quantenphysik geliehenen) Heisenbergschen Unschärferelation (auch als: Unbestimmtheitsrelation) (vgl. Machelett 1996: 1) ist nämlich eine gleichzeitige gute Auflösung von Frequenz und Zeit unmöglich. Doch bei der Dekodierung der Akustik als sprachliche Information zieht das Gehirn des Rezipienten die gesamte oszillographische Schwingungsstatistik und das gesamte „Rollspektrogramm“ höchstwahrscheinlich mit in Betracht, und es stellt mittels Transformationen und Durchschnitten Korrelationsfunktionen auf (vgl. Tarnóczy 1974: 181). Jedenfalls ist/wäre ein manueller Eingriff hoher Exaktheit in automatisierte Messvorgänge angebracht, um das Unschärfeproblem zumindest partiell zu beheben (und mithin der Erzeugung – Abstrahlung – Wahrnehmung/Dekodierung Korrelation näher zu kommen).
7. Mit Kenntnis der vorhandenen Messdaten ist lediglich die Formulierung von gewissen Korrelationsregeln des Paares Physiologie – Akustik möglich. Auch Valaczkai (1993: 166) gibt der Ansicht Ausdruck, dass „[...] es [...] bisher noch nicht gelungen ist, die Segmente der akustischen Registrate [...] den einzelnen Phasen der physiologischen Erzeugung des Sprachschalls restlos zuverlässig zuzuordnen“. Das ist ein weiteres Indiz dafür, dass sowohl die Technologie der Erschließung hochkomplexer Vorgänge bei der Erzeugung des Sprechschalls als auch die Messdaten weiter zu präzisieren wären.

## 1.2 Zielsetzung

Die vorliegende experimentalphonetische Untersuchung beabsichtigt für die oben angeschnittenen Probleme einen theoretisch-method(olog)ischen Lösungsvorschlag zu geben, und zwar am Beispiel des Schwas im Deutschen, mit dem Vorbehalt, dass wegen des limitierten Umfanges des Manuskriptes das Grundproblem (GP) 4 nicht behandelt werden kann. Der Beitrag setzt sich also zum Ziel, die zeitliche Variabilität (GP 5 – 6) der Komponenten der akustischen Struktur des Schwas im Deutschen (vgl. Valaczkai 1998: 78 u. die Bildtafel 23) in den Sequenzen  $V_{[ə]}C_{[s]}$  und  $V_{[ə]}C_{[ʃ]}$ , d.h. in der besonders problematischen oppositiven Lautverbindung [stimmhaft] – [stimmlos] möglichst genau, mit dem Einsatz neuer Methoden auszulegen. Bei den instrumentell durchgeführten

Ermittlungen der den materiell-energetischen Lautkörper bildenden Parameter der Sprechlaute werden die Assimilationserscheinungen in Lautübergängen zu [s] und [ʃ] (GP 3) freilich mit berücksichtigt. Weiterhin haben solche Faktoren wie Periodizität und Harmonizität u.a. eine hohe Priorität (GP 1 – 2), die in diesem Kontext noch nicht ausreichend beleuchtet und bisher lediglich von anderen, hauptsächlich medizinischen Wissenschaftszweigen, der Laryngologie und der Audiologie behandelt wurden. Es wird ferner angestrebt, durch minuziöse manuelle Messungen eine Korrelationsfunktion von den prozentualen Werten der Dehnung/Verkürzung glottaler Periodendauern und der Bewegung der Formantstruktur resp. deren Wahrnehmung auf der Basilmembran der Cochlea aufzustellen (GP 6 – 7). Die erstellten dynamischen Registrare werden bezüglich der numerischen Werte z.T. tabellarisch und verbal ausgewertet, z.T. wird auf das bewährte Verfahren der akustischen Phonetik, die Registrare als nonverbale, doch verbalisierbare visuelle Informationsquellen zu behandeln, zurückgegriffen.

### 1.3 Theoretische Erwägungen

Laut der aerodynamisch-myoelastischen Theorie der stimmhaften Lauterzeugung (vgl. Berg 1968, bes. die Seiten 291–301) sprengt der sich unterhalb der Stimmlippen bildende subglottale statische Luftdruck (Überdruck) den Glottisverschluss und presst hierauf den Luftstrom durch die düsenförmige Öffnung in das Ansatzrohr, wobei die Druckdifferenz unter- und oberhalb der Glottis jäh ausgeglichen wird. Die Luft strömt an den gewölbten Stimmlippen schnell vorbei und erzeugt – laut der Bernoullischen Strömungsgleichung – einen statischen Unterdruck, der die Stimmlippen abermals zusammenzieht. Die zyklische Wiederholung des Vorgangs zeitigt eine periodische Druckschwankung auf der Glottis, die auf das Ansatzrohr übergreift (vgl. Tillmann/Schiel 1995: 4, Nepert/Pétursson 1986: 95) und sich akustisch als ein im Zeitbereich periodisches Schallsignal manifestiert. Bei der stimmlosen Lauterzeugung verengt sich das Ansatzrohr irgendwo (bei [s] wird eine schmale Enge zwischen dem Apex/Prädorsum und den oberen Inzisiven gebildet, bei [ʃ] ebenso, doch zwischen dem Apex/Prädorsum und dem Zahndamm/Präpalatum) und es steigt um die und besonders in der Hemmstelle die Strömungsgeschwindigkeit. Beim Überschreiten des kritischen Wertes entsteht eine chaotisch-turbulente Luftverwirbelung, deren akustische Manifestation ein im Zeitbereich aperiodisches Schallsignal, ein Rauschen ist. Verengen kann sich das Ansatzrohr natürlich an mehreren Stellen; die Dimensionen bestimmen die geometrische Konfiguration des Resonanzraumes. Laut Tillmann (1980, nach Tillmann/Schiel 1995: 4) ist die Stärke der Verformung des Quellensignals durch die artikulatorische Tiefe, die Hinterverengung beeinflusst.

Akustisch setzt sich die stimmhafte Erzeugung des Sprechschalls aus zwei Teilmechanismen zusammen: aus der Rohschallerzeugung (Quelle) und einem

nachgeschalteten Resonator (Filter) (vgl. Fant 1968, bes. 191ff.). Durch die Quelle (Schwingung der Stimmlippen) wird ein (vielfrequentes und prinzipiell harmonisches) Schallsignal erzeugt, durch die erzwungene Mitschwingung unmittelbar über der Glottis gedämpft und die Resonanzräume des Ansatzrohres passierend bzw. durch sie verformt an den Mundlippen abgestrahlt. Bei der stimmlosen Erzeugung gibt es keine Stimmlippenschwingung: Als Quelle fungiert die chaotisch-turbulente Luftverwirbelung. Der Übergang aus dem einen (stimmhaften) in den anderen (stimmlosen) Zustand ist sowohl physiologisch als auch akustisch hochkomplex, seine Deskription ist daher oft problematisch. Ein weiteres erhebliches Problem besteht darin, dass die relevanten Merkmale der kombiniert verlaufenden Dekodierung akustischer Strukturvariablen als Phoneme und Morpheme im zeitlichen Kontinuum des Sprechschalls vermutlich in der Unterscheidung zwischen sich kontinuierlich ändernden Klang- und Geräuschqualitäten liegen. Dies geschieht theoretisch auf Grund eines stetigen Vorwärts- und Rückwärtsvergleichs, der sich nicht nur auf die (in gewissen Toleranzen) stationären Mittenphasen der Sprechlaute, sondern auch auf segmentkombinationsspezifische Qualitätsänderungen in den Lautübergängen erstreckt (vgl. Neppert 1997: 239).

## 1.4 Methode

### 1.4.1 Korpusauswahl, technische Parameter, die Analysesoftware

Die exakte Ermittlung der akustischen Gestalt sprechsprachlicher Äußerungen ist ohne das Vorliegen von Störsignalen lediglich in einem reflexionsfreien und schalltoten Raum (e.g. in einem Studio) möglich (vgl. Lindner 1969: 42). In einem solchen Raum kann die auditive Kontrolle des Sprechvorgangs durch den Produzenten, d.h. die Rückkopplung nur durch die Zwischenschaltung technischer Geräte erfolgen. Dies ist aber nicht die übliche Situation bei der alltäglichen Sprechkommunikation und es kann sich auf die natürliche Sprechproduktion ungünstig auswirken. Von diesen Umständen, hauptsächlich von dem Fehlen des Reflexionsschalles oder dessen technischem Ersatz können sich jedoch geschulte Rundfunksprecher unabhängig machen, die *ex offa* daran gewöhnt sind, ihre eigene Stimme durch Kopfhörer zum Hören zu bekommen. Sie bedienen sich ferner weder der durch Theatersäle bedingten kräftigeren Lautung und der überspannten Artikulation des „Siebs“ noch dialektaler Elemente. Vor diesen Hintergründen empfahl sich, das Korpus an einem (vor allem umfangreicheren) Hörfunkmaterial zu bestimmen. Frühere Experimente zeigten, dass die emotional gefärbte Rede nicht reale, akustisch „trügerische“ Ergebnisse übermitteln, so wurden emotional neutrale Tonmaterialien, Nachrichtensendungen,

nicht aber Debatten oder Reportagen verarbeitet. Die sorgfältige Auswahl der Versuchsperson(en) garantiert die Möglichkeit einer sprachlich validen Typusbildung: Es waren v.a. Geschlecht und Alter der VP in Betracht zu ziehen. Weibliche Sprecher haben eine akustisch schwer analysierbare Stimme, in erster Linie wegen der hochgradigen intonatorischen Variabilität. Der Faktor Alter ist für die „Klangqualität“ verantwortlich, man denke nur an die „stumpfen“ oder heiseren Altersstimmen. Also wurde ein männlicher Sprecher mittleren Alters in der Audio-Ausgabe der „Zeit“ ausgewählt, dessen Aussprache mit der im Wörterbuch der deutschen Aussprache (1974) kodifizierten einen stetigen Vergleich erfuhr. Obwohl sich die Untersuchung auf ein viel größeres Korpus (von etwa 13 Stunden) erstreckte, werden aus Gründen des Umfangs die akustischen Parameter lediglich zweier Schwa-Laute vorgeführt, die das Lautphänomen in den Sequenzen  $V_{[ə]}C_{[s]}$  und  $V_{[ə]}C_{[ʃ]}$  nach vorherigen akustischen Analysen zu repräsentieren fähig sind. Die Wörter, die die Laute beinhalten sind wie folgt: glückliches, gespannt. Technische Parameter: mp3-Kodierung, Datenfluss: 96 kbps<sup>2</sup>, Abtastfrequenz: 44100 Hz. Zur computerisierten Schallanalyse wurden die Daten in \*.wav PCM<sup>3</sup> konvertiert. Als Analysesoftware diente das als *Freeware* erhältliche Programmpaket PRAAT 4.3.04, das neben akustischen Analysen auch Segmentierung, Etikettierung u.a. ermöglicht. Bei der Kalibrierung der Software wurden stets die Anweisungen im Help-Menü befolgt.

#### 1.4.2 Katalog und terminologische Erklärungen der Untersuchungsfaktoren

1. Die relative/spezifische Bildungsdauer ist nicht nur lautkontextabhängig, sondern sie beeinflusst auch das in erster Linie vom Gefühlszustand dependierende Sprechtempo. Doch ist m.E., d.h. durch die oben angesprochene akkurate Korpusauswahl ein vermutlicher Mittelwert zu ermitteln.<sup>4</sup> Faktorzeichen (FZ):  $BD^{rs}$ . Maßeinheit (ME): ms (Millisekunde). Gemessen wird ferner die Dauer der sich als Übergang zu  $[s]/[ʃ]$  manifestierenden Assimilationszone. FZ: AZ. ME: ms. Die Auslenkungsamplitude (auch als: Elongation) ist der positive/negative Maximalausschlag einer Schwingung von der Ruhelage zu einem Zeitpunkt. FZ:  $A^-/A^+$ . Examiniert wird auch ihre temporale Anordnung im Verhältnis zu  $BD^{rs}$ ; die numerischen Werte werden verallgemeinert

---

<sup>2</sup> Kilobit pro Sekunde

<sup>3</sup> *pulse code modulation* (das analoge Signal wird binär kodiert)

<sup>4</sup> Die zu diesem Zweck bestens geeigneten statistischen Methoden würden den notgedrungenen Rahmen des Beitrags bei Weitem sprengen.

prozentual umgerechnet. Die RMS<sup>5</sup>-Amplitude ist die Wurzel aus dem durchschnittlichen Amplitudenquadrat. FZ:  $A^{\text{rms}}$ . Die ME für alle angeschnittenen Amplituden ist Pa (Pascal). Die Schallenergie (als Summe der potentiellen und kinetischen Energie, d.i. das im System gespeicherte Arbeitspotential) in der Luft, FZ:  $E^{\text{idL}}$ , wird in  $\text{J}/\text{m}^2$  (Joule pro Quadratmeter) angegeben. Die Schallintensität (auch als: Schallstärke), FZ:  $I$ , ist „[...] als die Energiemenge definiert, die pro Sekunde auf einer senkrecht zur Ausbreitungsrichtung befindlichen Fläche von  $1 \text{ m}^2$  ankommt [...]“ (Neppert/Pétursson 1986: 57). Demgemäß wird sie in  $\text{W}/\text{m}^2$  (Watt pro Quadratmeter) gemessen. Beim absoluten Schalleistungspegel (in der Fachliteratur meist fälschlich als Schalldruckpegel) liegt der Zeropunkt bei  $2 \cdot 10^5 \text{ N}/\text{m}^2$  (Newton pro Quadratmeter) Wechseldruck, d.h. der unteren Hörschwelle<sup>6</sup>. FZ: P·ME: dB SPL<sup>7</sup> (Dezibel SPL).

2. Unter Jitter versteht man die Mikrovariationen in der Periodizität der Stimmlippenvibration und ihrer akustischen Manifestation, der Grundfrequenz. Mit Shimmer wird die Mikrovariation in der Amplitude bezeichnet. In Anlehnung an Titze (1988, 1994, 1995) erwähnt auch Freenstein (2000: 8) das Phänomen:

Um einen Nettoenergieübertrag aus der kinetischen Energie der Strömung auf die Schwingung zu gewährleisten, muss der intraglottische [sic!] zeitliche Druckverlauf zwischen zwei Verschlüssen eine Asymmetrie aufweisen. Diese Asymmetrie kann durch zwei verschiedene Mechanismen hervorgerufen werden: Die Trägheit der bewegten Luftsäule und nichtgleichmäßige Schwingungsmoden. Die nichtgleichmäßige Schwingungsmoden entsteht durch [sic!] eine Überlagerung der Schwingung des Körpers der Stimmlippe (body) mit einer Welle, die in der Schleimhaut läuft, welche die Stimmlippe überzieht (cover).

Der Natur von relativ kurzer Bildung und Dauer der Sprechlaute resp. dem Miteinbezug der stark geräuschhaften Übergänge zu [s]/[ʃ] gemäß der Analyse werden die Messungen an der Kurzzeitfrequenz- und Kurzzeitamplitudenperturbation vorgenommen. Die prozentualen Werte repräsentieren die Standardabweichungen der Grundfrequenz dividiert durch den mittleren Grundfrequenzwert (Jitter) resp. die Standardabweichungen der Amplitude dividiert durch den mittleren Amplitudenwert (Shimmer). Jitter (local) ist der Mittelwert der Absolutdifferenz konsekutiver Periodendauern dividiert durch die mittlere Periodendauer.

---

<sup>5</sup> *root mean square*

<sup>6</sup> ~ richtet sich nach  $2 \cdot 10^5 \text{ Pa}$  (Referenzdruck), d.i. die normative untere Hörschwelle für eine Sinuswelle von 1000 Hz. Normativ in dem Sinne, dass sie mittels empirisch-statistischer Untersuchungen an einer großen Zahl von jugendlichen Normalhörern ermittelt bzw. festgesetzt worden ist (vgl. Neppert/Pétursson 1986: 59).

<sup>7</sup> *sound pressure level*

er. FZ: Jitt. ME: %. Jitter (local, absolute) ist der Mittelwert der Absolutdifferenz konsekutiver Periodendauern. FZ: Jita. ME: s. Jitter (rap)<sup>8</sup>: Jitt geglättet über 3 Perioden. FZ: J<sup>rap</sup>. ME: %. Jitter (ppq5)<sup>9</sup>: Jitt geglättet über 5 Perioden. FZ: J<sup>ppq5</sup>. ME: %. Shimmer (local) ist der Mittelwert der Absolutdifferenz konsekutiver Periodenamplituden dividiert durch die mittlere Amplitude. FZ: Shim. ME: %. Shimmer (local, dB) ist der Mittelwert der Absolutdifferenz konsekutiver Periodenamplituden. FZ: ShdB. ME: dB. Shimmer (apq3): Shim geglättet über 3 Perioden. FZ: S<sup>ppq3</sup>. ME: %. Shimmer(apq5): Shim geglättet über 5 Perioden. FZ: S<sup>ppq5</sup>. ME: %. Diese Werte werden maschinell errechnet!

3. Unter Harmonizität wird herkömmlich die Energieverteilung zwischen den harmonischen und den Geräuschkomponenten des Sprechschalls verstanden. Als Nullpunkt, 0 dB versteht man die gleichmäßige Energieverteilung zwischen den beiden. Aufgeführt wird hier lediglich das Maximum; weitere Daten sind aus der entsprechenden Abbildung (Abb. 2) unschwer zu entnehmen. FZ: HNR. ME: dB.
4. Die (sprech- und sprachlautspezifische) inhärente Grundfrequenz ist physiologisch mit der sekundlichen Häufigkeit der Stimmlippenvibration bei der Erzeugung stimmhafter Laute, akustisch mit dem Frequenzwert des ersten und über die größte Amplitude verfügenden Obertons des Stimmklanges identisch. Das Spezifikum ist den durch diverse Zungenstellungen bewirkten Spannungsänderungen in der laryngalen Muskulatur resp. der je nach Sprechlaut unterschiedlichen Luftströmung bei der Erzeugung beizumessen (vgl. Neppert/Pétursson 1986: 149ff.). Da sich sowohl die Grundfrequenz als auch deren Amplitude ebenfalls intersegmental (sprechlautintern) variieren können, wird lediglich der Wert mit der größten Amplitude aufgeführt; weitere Daten sind aus der entsprechenden Abbildung (Abb. 2) unschwer zu entnehmen. FZ:  $F_0$ . ME: Hz (Hertz). FZ ihrer Amplitude:  $A_0$ . ME: dB (nicht aber in Pa od. Pa/Hz!).
5. Die sich ausbreitende Schallwelle bei der Erzeugung wird an den *in theoria* schallharten Wänden des Ansatzrohres reflektiert und wegen der Linearität des Systems kommt es zu Überlagerungen und Additionen der Reflexionen (Superpositionsprinzip). Die intensivsten Eigenresonanzen der Schallwelle in der Hohlraumgestaltung des Ansatzrohres geben der Klangqualität des Sprechlautes ein starkes Gepräge und nennen sich Formanten. Die ersten zwei sind nicht sprecherspezifisch (vgl. Kassai 1998: 72), sondern sie tragen allgemeine sprachliche Informati-

---

<sup>8</sup> *relative average perturbation*

<sup>9</sup> *period perturbation quotient*

onen. Demgemäß werden nur sie behandelt. Sie werden manuell, anhand von Breitbandspektrogrammen und Spektren ermittelt: Die  $BD^{rs}$  wird auf glottale Perioden aufgeteilt, deren nacheinander gestellte und um  $90^\circ$  nach links gedrehte Spektren die Amplitudengestaltung der Formanten in der Zeitfolge zeigt. Es wird lediglich der Frequenzwert mit der größten Amplitude aufgeführt; weitere Daten sind aus der entsprechenden Abbildung (Abb. 3 u. 3.1 bzw. 3.2) unschwer zu entnehmen. Examiniert wird auch ihr Verhältnis zu  $F_0$  auf der Frequenzachse; die Ergebnisse werden verallgemeinernd abgerundet. FZ der Formanten:  $F_1$  bzw.  $F_2$ . ME: Hz. FZ der Amplituden:  $A_1$  bzw.  $A_2$ . ME: dB. Da das Ansatzrohr ein reales physikalisches System ist, ist es verlustbehaftet, d.h., dass ein Energieaustausch zwischen ihm und seiner Umwelt stattfindet. Die Verluste, hauptsächlich infolge Reibung der Luftmoleküle aneinander und dem Mitschwingen (der in Wirklichkeit nicht idealharten) Wände entziehen dem System Energie: Die Schwingungen werden in der Zeitfolge gedämpft, die Bandbreiten werden im Frequenzbereich hingegen vergrößert. Gemäß dieser umgekehrten Proportionalität wird eine Frequenzbandbreite der Formanten angenommen, die sich „[...] als der Frequenzbereich [...] [definiert], den der Formant bei -3 dB unter seiner Spitzenamplitude in der Breite einnimmt“ (Pompino-Marschall 2003: 109). FZ der unteren und der oberen Grenze der Frequenzbandbreite der Formanten:  $B_1^u - B_1^o$  bzw.  $B_2^u - B_2^o$ . ME: Hz. Weitere Variablen bei der Bestimmung des sprachlich informativen und daher in Bezug auf die Dekodierung relevanten Teils der Formantstruktur sind die Energie der Bandbreiten, FZ:  $EB_1$  bzw.  $EB_2$ , ME:  $Pa^2s$ , und die Dichte der Bandbreiten, FZ:  $DB_1$  bzw.  $DB_2$ , ME:  $Pa^2/Hz^2$ .

6. Die Reaktionsareale und ihre psychoakustischen Indizien auf der Basilar membran der Cochlea (Schnecke) bei der Wahrnehmung werden mittels Cochleogramm modelliert, doch die primären Daten, die numerischen/prozentualen Werte leistet der aus dem Spektrum hergeleitete Stimuligraph. Der Grund dafür ist, dass das Ohr als ein perzeptorisches Instrument für Schall die zu ihm kommenden zusammengesetzten Schwingungen in Sinuskomponenten zerlegt (aus denen ein sog. „Neuralspektrogramm“ herauswächst, vgl. Kassai 1998: 86), etwa so wie das mathematische Gegenstück dieses Phänomens, die Fouriertransformation, die als Basis für das Spektrum fungiert. Die wahrgenommene Tonhöhe, FZ: Z, wird in Bark<sup>10</sup>, der „[psycho]akustische[n] Maßeinheit

---

<sup>10</sup> Die Bark-Skala liefert die Korrelationsfunktion der subjektiv empfundenen Tonhöhe und der objektiven Frequenz.



der wahrgenommenen Tonhöhe von der Breite einer Frequenzgruppe<sup>11</sup> (Pompino-Marschall 2003: 305) angegeben. Der Lautstärkepegel (d.i. die wahrgenommene Lautstärke), FZ: LSP, wird in phon gemessen. Die Lautheit, FZ: LH, wird in sone<sup>12</sup> angegeben.<sup>13</sup> In Kenntnis der unter 2.1 aufgeführten Informationen zur Verbalisierung der zu diesem Punkt verfertigten Registrate (Abb. 4, 4.1, 4.2, 4.3, 4.4) kann die Interpretation dem Leser überlassen werden.

## 2 Durchführung der Untersuchungen

### 2.1 Registrate und Messtabellen

Zum eindeutigen Verständnis genügt die Information, dass all die hier aufgeführten Registrate kartesische Koordinatensysteme sind. Numerische Informationen sind von den Ordinaten (O) und den Abszissen (Ab) ablesbar. Erstes Registrat (Abb. 1): Oszillogramm (d.i. die vollständige Schwingungskurve) (O links) + Schalleistungspegel (O rechts) in der Zeitfolge (Ab). Zweites Registrat (Abb. 2): Harmonizität (O links) + inhärente Grundfrequenz (O rechts) in der Zeitfolge (Ab). Drittes Registrat (Abb 3): Spektrogramm. O: Frequenz (die waagrecht liegenden dunklen Balken sind die Formanten!), Ab: Zeitfolge. Aus dem Schwärzungsgrad, der quasi-räumlichen Dimension, der Applikate (Ap) ist auf die Amplitude der Formanten schlusszufolgern: Je schwärzer die Balken sind, desto größer ist die Amplitude der Formanten. Viertes Registrat (Abb. 3.1): Frequenzwerte der Formanten (O) in der Zeitfolge (Ab). Mit Strichellinie sind die Grenzen der Bandbreiten gekennzeichnet. Fünftes Registrat (Abb 3.2): Amplitudenwerte der Formanten (O) in der Zeitfolge (Ab). Sechstes Registrat (Abb. 4): Cochleagramm. O: Reaktionsareale auf der Basilarmembran der Cochlea (O) bei der Wahrnehmung des Sprechschalls in der Zeitfolge (Ab). Aus Ap ist auf den Lautstärkepegel zu folgen. Siebtes Registrat (Abb. 4.1): wahrgenommene Tonhöhe (als Z) (O) in der Zeitfolge (Ab). Achtes Registrat (Abb. 4.2): Lautstärkepegel (als LSP) (O) in der Zeitfolge (Ab.). Neuntes Registrat (Abb. 4.3): Lautheit (als LH) (O) in der Zeitfolge (Ab). Die Zeitachse wird je auf glottale Perioden (GP) aufgeteilt, die einer unglücklichen graphischen Limitation zufolge in den Registraten nur in stilisierter Form, als gleichmäßige Zeitsequenzen erscheinen. Die akustischen Tendenzen aber sind gut zu beobachten.

---

<sup>11</sup> Frequenzgruppe nennt man diejenige Bandbreite im auditorischen System, deren Signale zu einem Erregungspegel gebündelt verarbeitet wird.

<sup>12</sup> Sone ist die Maßeinheit der Verdoppelung der Lautheit. 1 sone entspricht den Parametern 1 kHz/40 dB/40 phon eines Referenztons.

<sup>13</sup> Die Wahrnehmungsgrößenskalen sind immer anhand einer großen Anzahl von mono- und binauralen Tonhöhevvergleichsexperimenten an jugendlichen Normalhörern rektifiziert.

[ə] in glückliches

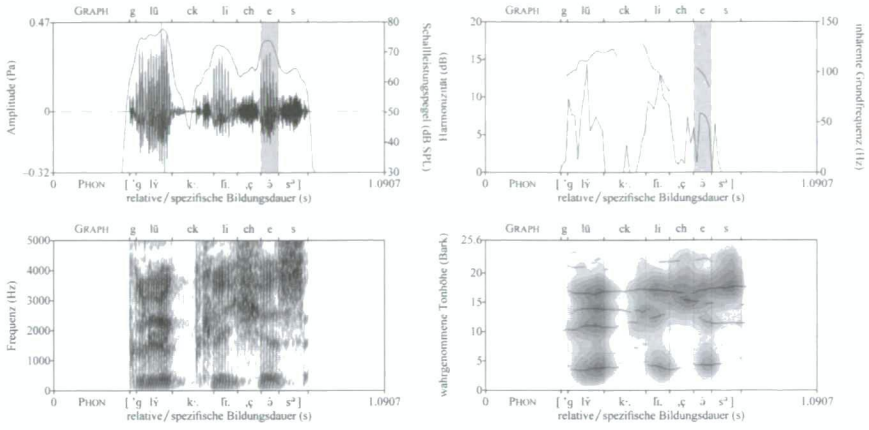


Abb. 1, 2, 3, 4

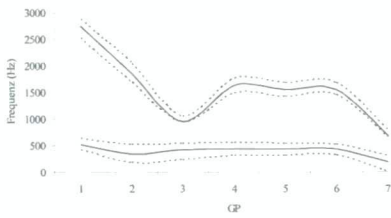


Abb. 3.1

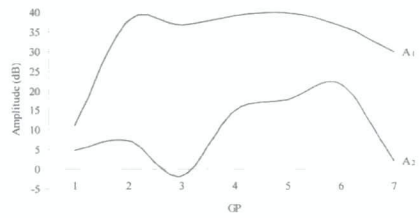


Abb. 3.2

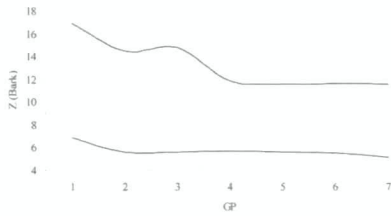


Abb. 4.1

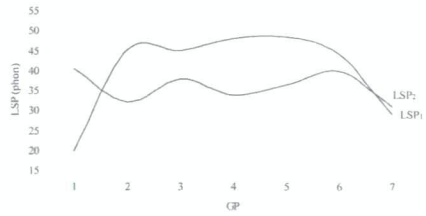


Abb. 4.2

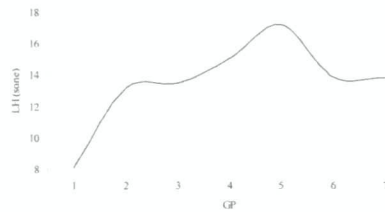


Abb. 4.3

[ə] gespannt

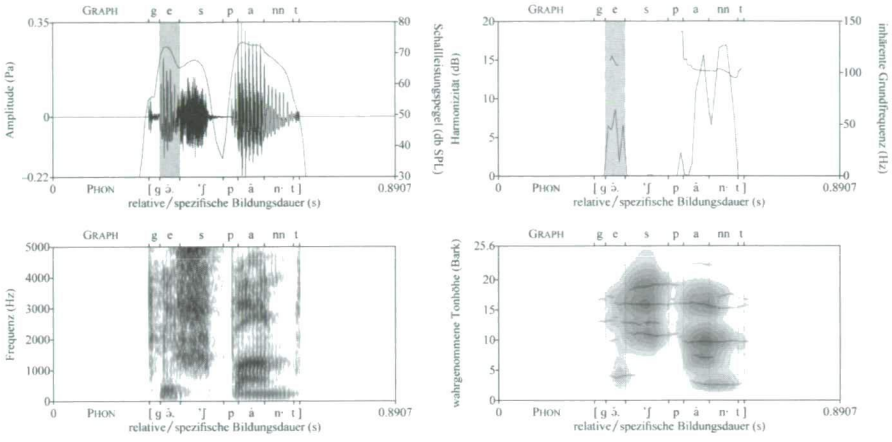


Abb. 1, 2, 3, 4

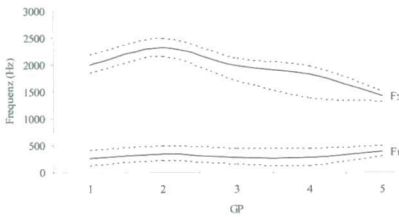


Abb. 3.1

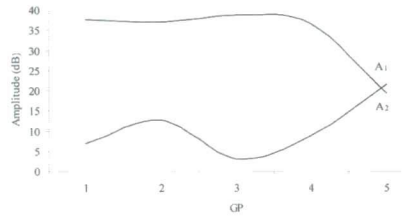


Abb. 3.2

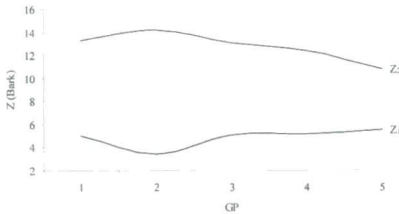


Abb. 4.1

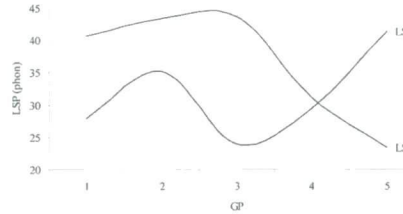


Abb. 4.2

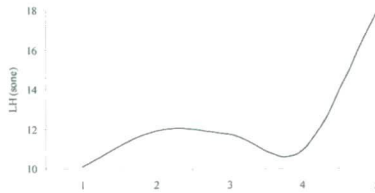


Abb. 4.3

## [ə] in glückliches

1.  $BD^{rs}$  ist 73.08 ms, 17.88 ms (etwa 24.46%) davon ist AZ.  $A^-$  liegt mit  $-0.17$  Pa bei 38.43 ms (etwa 52.58% der  $BD^{rs}$ ).  $A^+$  liegt mit 0.31 Pa bei 37.18 ms (etwa 50.87% der  $BD^{rs}$ ).  $A^{rms}$  ist 0.08 Pa.  $E^{idl}$  ist  $1.1783544e-06$  J/m<sup>2</sup>. I liegt bei  $1.612331e-05$  W/m<sup>2</sup>. P hat sein Maximum bei 73.85 dB SPL, in der AZ ist dieser Wert um 3.13 dB SPL weniger.

## 2. Jitter und Shimmer:

Jitt	7.126%
Jita	688.386E-6 s
J <sup>rap</sup>	1.773%
J <sup>ppq5</sup>	undef.
Shim	4.875%
ShdB	0.417 dB
S <sup>ppq3</sup>	3.250%
S <sup>ppq5</sup>	undef.

## 3–4. Harmonizität und inhärente Grundfrequenz:

Das Maximum der HNR-Kurve bildet sich bei 7.8 dB. In der AZ ist ein steiler und sehr intensiver Fall zu beobachten.

$F_0$  liegt bei 96.5 Hz,  $A_0$  ist 38.2 dB.  $F_0$  sinkt stetig.

## 5. Formantstruktur:

$F_1$ (Hz)	440
im Verh. zu $F_0$	5
$AB_1$ (dB)	39.7
$B_1^u$ (Hz)	0
$B_1^o$ (Hz)	629
$EB_1$ (Pa <sup>2</sup> s)	$1.33 \cdot 10^{-4}$
$DB_1$ (Pa <sup>2</sup> /Hz <sup>2</sup> )	$2.12e-07$
$F_2$	1553
im Verh. zu $F_0$	16
$A_2$	21.7
$B_2^u$	656
$B_2^o$	2873
$EB_2$	$1.87e-06$
$DB_2$	$8.43e-10$

## [ə] in gespannt

1.  $BD^{rs}$  ist 52.44 ms, 18.94 ms (etwa 36.11%) davon ist AZ.  $A^-$  liegt mit  $-0.15$  Pa bei 11.45 ms (etwa 21.83% der  $BD^{rs}$ ).  $A^+$  liegt mit 0.22 Pa bei 9.89 ms (18.85% der  $BD^{rs}$ ).  $A^{rms}$  ist 0.06 Pa.  $E^{idl}$  ist  $5.8337061e-07$  J/m<sup>2</sup>. I liegt bei  $1.112263e-05$  W/m<sup>2</sup>. P hat sein Maximum bei 71.86 dB SPL, in der AZ ist dieser Wert um 5.79 dB SPL weniger.

## 2. Jitter und Shimmer:

Jitt	12.094%
Jita	1160.865E-6 s
J <sup>rap</sup>	3.027%
J <sup>ppq5</sup>	undef.
Shim	1.699%
ShdB	0.146 dB
S <sup>ppq3</sup>	undef.
S <sup>ppq5</sup>	undef.

## 3–4. Harmonizität und inhärente Grundfrequenz:

Das Maximum der HNR-Kurve bildet sich bei 8.6 dB. In der AZ ist ein steiler und sehr intensiver Fall zu beobachten.

$F_0$  liegt bei 109 Hz,  $A_0$  ist 37.2 dB. Nach einem kurzen Antieg sinkt  $F_0$  stetig.

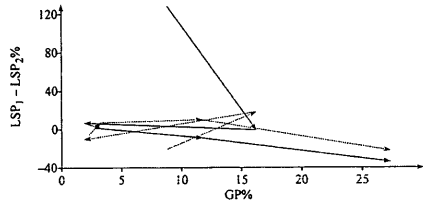
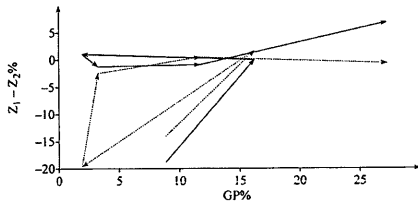
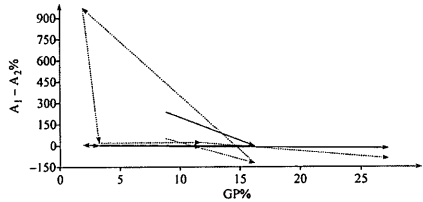
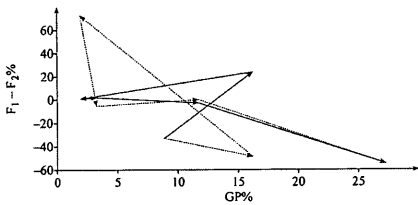
## 5. Formantstruktur:

$F_1$ (Hz)	268
im Verh. zu $F_0$	2
$A_1$ (dB)	38.7
$B_1^u$ (Hz)	118
$B_1^o$ (Hz)	283
$EB_1$ (Pa <sup>2</sup> s)	$1.51e-05$
$DB_1$ (Pa <sup>2</sup> /Hz <sup>2</sup> )	$9.17e-08$
$F_2$	1420
im Verh. zu $FB_0$	13
$A_2$	21.5
$B_2^u$	1312
$B_2^o$	2482
$EB_2$	$2.36e-07$
$DB_2$	$2.02e-10$

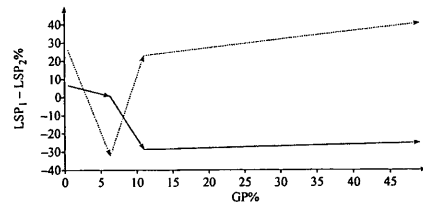
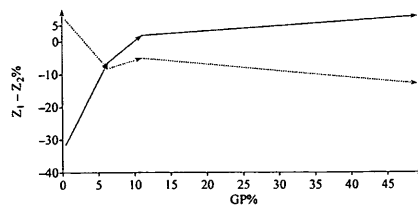
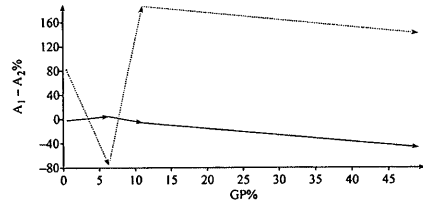
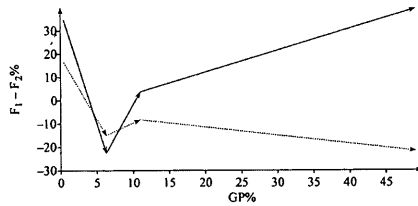
## 2.2 Korrelationsfunktionen

Die numerischen Werte der Dehnung/Verkürzung glottaler Periodendauern und der Bewegung der Formantstruktur resp. deren Wahrnehmung auf der cochlearen Basilmembran werden prozentual umgerechnet und in Korrelationsfunktionen dargestellt. Diese Methode eignet sich gut zur Beobachtung von Zusammenhängen physiologischer und (psycho)akustischer Vorgänge bei der Erzeugung des Sprechschalls und dessen Verarbeitung im auditorischen System. (Die Strichlinie bezeichnet F, A, Z, LSP<sub>2</sub>.)

### [ə] in glücklichem



### [ə] in gespannt



### 3 Fazit

Das über keine autonome Bedeutung verfügende, jedoch im syntagmatischen Verlauf der Rede schon ausgelegte Segment, der Sprechlaut ist in der Zeitfolge innerlich-akustisch vielfach fein strukturiert. Die Dynamik der akustischen Information, die die Möglichkeit der Dekodierung als sprachliche Information beinhaltet und der als Bauelement eine komplexe Bedeutung zukommt, blieb bis auf wenige Ausnahmen unerschlossen. In diesem Beitrag wurde der Versuch unternommen, die zeitliche Variabilität der Komponenten der segmentalen akustischen Struktur anhand dynamisch fundierter Forschungsmethoden auszulegen, und zwar am Beispiel des Schwas im Deutschen in den Sequenzen  $V_{[ə]}C_{[s]}$  und  $V_{[ə]}C_{[ʃ]}$ . Erzielt wurde ferner die Aufstellung von Korrelationsfunktionen der Physiologie der Erzeugung und der Akustik der Abstrahlung resp. deren Verarbeitung im auditorischen System des Rezipienten. Die Methode und die durch sie gewonnenen Ergebnisse lassen sich bei der Neudefinierung der Akustik der segmentalen Struktur des Sprechschalls, ferner bei der Erarbeitung einer akustischen Phonologie für das Deutsche gut verwenden.

### Bibliographie

- Berg, J.W. van den: Mechanism of the larynx and the laryngeal vibrations. In: Malmberg, Bertil: Manual of Phonetics. Completely revised and extended edition of Manual of Phonetics, edited by L. Kaiser, published in 1957. Amsterdam 1968, 278–308.
- Bolla, Kálmán: A Conspectus of Russian Speech Sounds/Атлас звуков русской речи. Budapest/Wien 1981 (Slavistische [sic!] Forschungen, 32).
- Bolla, Kálmán/Valaczkai, László: Német beszédhangok atlasza. Budapest 1986 (Magyar Fonetikai Füzetek/Hungarian Papers in Phonetics, 16).
- Fant, Gunnar: Analysis and synthesis of speech processes. In: Malmberg, Bertil: Manual of Phonetics. Completely revised and extended edition of Manual of Phonetics, edited by L. Kaiser, published in 1957. Amsterdam 1968, 173–277.
- Freienstein, Heiko: Vokaltraktmodellbasierte Schätzung von Steuerparametern eines Moduls zur Sprechernormalisierung. Dissertation. Göttingen 2000. Auch im Internet unter: <http://webdoc.sub.gwdg.de/diss/2000/freienstein/freienstein.zip> [gesehen am 04.05.2005].
- Gósy, Mária: Szükséges és szükségtelen hangátmenetek. In: Gósy, Mária (Hrsg.): Beszédkutatás. A Beszédkutatás '95 konferencia válogatott tanulmányai. Budapest 1995, 20–31.
- Kassai, Ilona: Fonetika. Budapest 1998.
- Lindner, Gerhart: Einführung in die experimentelle Phonetik. Berlin 1969.

- Machelett, Kirsten: Das Lesen von Sonagrammen. Begleitendes Hypertext-Dokument zur Vorlesung. München 1996. Im Internet unter: <http://www.phonetik.uni-muenchen.de/SGL/SGLHome.html> [gesehen am 04.05.2005].
- Neppert, Joachim/Pétursson, Magnús: Elemente einer akustischen Phonetik. 2., überarb. u. erw. Aufl. Hamburg 1986.
- Neppert, Joachim: Zur Relevanz von akustisch-auditiv-phonetischer und audiologischer Terminologie und von exakter Meßwerterfassung in diesem Bereich für die Hörgeschädigtenpädagogik. In: Frerichs, Hajo/Günther, Klaus-B./Neppert, Joachim (Hrsg.): Perspektiven in der Schwerhörigenpädagogik. Festschrift zum 65. Geburtstag von W. Hartwig Claußen. Heidelberg 1997, 229–250.
- Pompino-Marschall, Bernd: Einführung in die Phonetik. 2. Aufl. Berlin, New York 2003.
- Tarnóczy, Tamás: A magánhangzók akusztikai vizsgálatának problémái. In: Tegledi, Zsigmond/Szépe, György (Hrsg.): Általános nyelvészeti tanulmányok X. A nyelv hangdomíniuma. Budapest 1974, 181–196.
- Tillmann, Hans/Schiel, Florian: Akustische Phonetik. Begleitendes Hypertext-Dokument zur Vorlesung. München 1995. Im Internet unter: <http://www.phonetik.uni-muenchen.de/AP/APHome.html> [gesehen am 06.05.2005].
- Valaczkai, László: Zur komplexen physiologischen und akustischen Typologie der deutschen Verschlusslaute. In: Földes, Csaba (Hrsg.): Germanistik und Deutschlehrerausbildung. Festschrift zum hundertsten Jahrestag der Gründung des Lehrstuhls für deutsche Sprache und Literatur an der Pädagogischen Hochschule Szeged. Szeged, Wien 1993, 165–181.
- Valaczkai, László: Atlas deutscher Sprachlaute. Instrumentalphonetische Untersuchung der Realisierung deutscher Phoneme als Sprechlaute. Wien 1998.
- Vértes O., András: Az artikuláció akusztikus vetülete. In: Bolla, Kálmán (Hrsg.): Fejezetek a magyar leíró hangtanból. Budapest 1982, 155–163.
- Wängler, Hans-Heinrich: Atlas deutscher Sprachlaute. 7., unveränd. Aufl. Berlin 1981.
- WddA = Krech, Eva-Maria/Kurka, Eduard/Stelzig, Helmut u.a. (Hrsg.): Wörterbuch der deutschen Aussprache. Leipzig 1974.