

Mária Gósy

DIE VERÄNDERUNGEN DER VOKALQUALITÄT IN SPONTANER REDE

EINLEITUNG

Die Untersuchung der akustischen Vokalstruktur blickt auf eine bereits mehrere Jahrzehnte alte Vergangenheit zurück. Aufgrund methodologischer Eigenheiten bezogen sich die Analysen eher auf gelesene Textdetails, zumeist auf die Vokale isolierter Wörter (Tarnóczy 1941; Molnár 1970; Bolla 1982). Mit dem Aufkommen der Sprachsynthese wurde eine Art von Kontrolle möglich, die sich auf die Richtigkeit bzw. Korrektur der Analysedaten bezog (Olaszy 1985). Es geschahen auch einige Versuche hinsichtlich des Zusammenhanges der Qualität der natürlich ausgesprochenen Vokale und der akustischen Struktur (Tarnóczy 1965; Gósy 1992; Szalai 1995). Dagegen gibt es keine zuverlässigen Angaben über die Beziehung der Vokalqualität der Wörter in spontaner Rede und der akustischen Struktur. Die ungarische Sprache reduziert – als sprachspezifische Erscheinung – die Vokale nicht wie das Englische (Jones 1966) oder Deutsche (Lindner 1981: 238). Die Reduzierung der Vokale in den genannten Sprachen bedeutet, daß eine gegebene Lautqualität als neutraler Vokal, Schwa, erscheint, entsprechend der sprachlichen Regeln, üblicherweise als Realisation des Vokalphonems einer unbetonten Silbe. Das Schwa ist also in jedem Fall die reduzierte Variante eines Phonems, im Holländischen erscheint es z.B. immer mit kürzerer Dauer als die kurzen Vokale (Koopmans-van Beinum 1994). Das Schwa erwähnt Laziczius als zentralen Vokal (vgl. 1963: 73).

Die Zeitverhältnisse von Wörtern in spontaner Rede sind selbstverständlich nicht mit denen isoliert ausgesprochener Wörter identisch, infolgedessen erscheint der Dauerunterschied in den die Wörter bildenden Sprechlauten. In fließender Rede erscheint auch im Ungarischen eine Art natürlicher Verkürzung, vor allem beim Aussprechen der Vokale, es gibt aber keine zuverlässigen Angaben über ihre systematische Analyse. Diese Art Verkürzung bedeutet theoretisch nicht zugleich eine qualitative Veränderung; die ungarischen Vokale werden unabhängig von ihrer Stellung in der Lautreihe und ihrem unmittelbaren Kontext immer vollwertig gebildet, es gibt also im Ungarischen keine qualitativ reduzierten Vokale als reguläre Realisationen gewisser Phoneme. Wir wollen nun von diesen Schwa-Erscheinungen einmal absehen, welche üblicherweise zwischen zwei Konsonanten, als physiologischer Zwang oder Marker von Wortgrenzen erscheinen (vgl. Kassai 1978). In diesen Fällen wird nämlich ein tatsächlich "neutrales" vokalartiges Lautdetail in unterschiedlicher Funktion in die laufende Rede eingeschaltet.

Gleichwohl wirft eine Beschleunigung des Sprechtempos die Möglichkeit auf, ob nicht gegen Ende des 20. Jahrhunderts auch in den Wörtern spontaner ungarischer Rede die reduzierten Vokale auftraten, die sich also hinsichtlich ihrer Lautqualität nicht mehr oder kaum noch voneinander unterscheiden. Die Frage ist, wie z.B. die Vokale des Satzes "*Kitaláltatok az eredetit?*" in spontaner Rede erscheinen können. Folgende Alternativen gibt es (in IPA-Transkription):

- a) [kit0la:lta:tok 0z eredetit]
- b) [kit0la:lta:t«k «z eredetit]
- c) [kit«la:lt«t«k «z er«d«tit]

Der erste Fall entspricht den sprachspezifischen Erwartungen; unabhängig von der aktuellen Dauer entspricht die Qualität der ausgesprochenen Vokale der der geplanten Vokale. Im zweiten Fall kommt es zu einer teilweisen Qualitätsreduktion; ein Teil der Vokale in unbetonter Stellung schlagen qualitativ in einen anderen Laut um, der ausgesprochene Vokal entspricht also nicht dem geplanten Sprechlaut. Im dritten Fall verliert – theoretisch – jeder in möglicherweise reduzierbarer Position befindliche Sprechlaut seine ursprüngliche (geplante) Qualität. Die Hypothese lautet, daß die Qualität der Vokale in betonter Stellung beständig ist, also der realisierte Sprechlaut jederzeit dem geplanten Sprechlaut entspricht.

Beim Hören spontaner Rede hat der Sprecher/Hörer – fast unabhängig vom durchschnittlichen Sprechtempo – immer den Eindruck, vollwertig gebildete Vokale zu hören. Die eventuellen Reduktionen fallen nur dann auf (und auch dann nicht in jedem Fall), wenn sie eine ganze Silbe betreffen oder mit dem völligen Ausfallen eines Lautes verbunden sind. Die Erklärung dafür liegt im Wirken des Redeverständnis-Mechanismus. Im mit den Redewahrnehmungsebenen auf quasi-parallele Weise vor sich gehenden Prozeß der Worterkennung werden die Entscheidungsergebnisse der unteren Ebenen von der/den oberen Ebene/n revidiert, was dann das Erkennen der der gegebenen Semantik entsprechenden Lautreihe bewirkt wird (Gósy 1989). In einer normalen Kommunikationssituation ist der Kontexteinfluß für das Sprechverständnis bestimmend, die Funktion der akustischen und phonetischen Ebene der Redewahrnehmung läßt sich nicht beurteilen, da der Mechanismus fast sofort auf die Ebene der semantischen und morphologisch/morphophonologischen Analysen "heraufschaltet". Auf den obigen Beispielsatz zurückkehrend, wird der Hörer auch im angenommenen Fall (c) das [«] in der zweiten und letzten Silbe nicht wahrnehmen, weil er als Ergebnis des als real-time vor sich gehenden Prozesses der Worterkennung die Bedeutung des Wortes *kitaláltatok* identifiziert, in dem die geplanten Vokale [0] und [o] unabhängig von ihrer aktuellen Realisierung in den betreffenden Silben vorkommen.

Die Hypothese ist, daß sich infolge der Beschleunigung des Sprechtempos sowie der Artikulationseigenarten der spontanen Rede auch im Ungarischen die Vokale reduzieren, sich diese Reduktion aber nicht auf die Zeitdauer beschränkt, sondern auch die Vokalqualität betrifft. Es wurde eine in höherem Maße neutrale Bildung der Vokale in spontaner Rede als in isolierten Wörtern bewiesen. Um diese Erscheinung eingehend akustisch und perzeptiv untersuchen zu können, wurde ein Versuch ausgearbeitet.

MATERIAL UND METHODE

Spontanes Material mit unterschiedlichen Themen eines jungen Erwachsenen ungarischer Muttersprache (Sprecher ungarischer Muttersprache, der nicht längere Zeit im Ausland lebte) wurde aufgenommen. Das durchschnittliche Artikulationstempo des Gewährsmannes betrug 12,75 Sprechlaute/s und die Minimal- und Maximalwerte 11,2 und 14,3 Sprechlaute/s. Entsprechend diesen Werten ist festzustellen, daß das Sprechtempo des ausgewählten Gewährsmannes mittelschnell und auch sein Artikulationstempo durchschnittlich (weder schnell noch langsam) zu nennen ist. Die Zeitdauer des gesamten Tontmaterials betrug 20 min, wovon für die vorliegende Untersuchung ein etwa 6 minütiger Teil bearbeitet wurde. In der ersten Untersuchungsserie wurden neben gleichzeitiger auditiver und visueller Rückkopplung (und Verstärkung) sämtliche Vokale der in spontaner Rede vorkommenden Wörter segmentiert, unabhängig von ihrer Stellung in der Lautreihe bzw. von ihren Betonungsverhältnissen. Bei der Segmentierung wurde die Bestimmung der sog. reinen Phase der Vokale angestrebt, es wurden also die Übergangsteile weggelassen. Nach der Segmentierung wurden die Dauer der Vokale (in ms) und der Wert der ersten drei Formanten (in Hz) bestimmt. Die akustischen Analysen wurden mittels unterschiedlicher Einstellungen eines Computerized Speech Lab, Modell 4300B durchgeführt.

In der zweiten Versuchsserie wurden fünf von den segmentierten Vokalen ausgewählt und in Random-Reihenfolge mit zweifacher Wiederholung auf ein Magnetophonband aufgenommen. Die ausgewählten Vokale sind folgende: von den Velaren [a:, 0, o], von den Palatalen [e, ø, ø:]. Die Auswahl war auf die Veränderung der Lautqualität gerichtet. Die dem reduzierten Vokal [«] am nächsten stehenden Laute [ø, ø:] fungierten gleichsam als "Etalon"; im analysierten Lautmaterial kamen die phonologisch kurzen und langen labialen Palatallaute mit fast identischem Anteil vor. Aufgrund des Abhörens der segmentierten Sprechlaute wurden jene Vokaltypen ausgewählt, die sich oftmals im neutralen Vokal realisierten. Im untersuchten Korpus realisierten sich die illabialen Palatalvokale oberer und mittlerer Zungenstellung kein einziges Mal als Schwa, ähnlicherweise auch der labiale Velarvokal oberer Zungenstellung nicht. Für die Analyse des labialen Palatalvokals oberer Zungenstellung war – obwohl seine Qualität mehrmals ins Schwa umschlug – aufgrund der

außerordentlich geringen Zahl dieser Laute keine Möglichkeit gegeben. Dementsprechend wurden sämtliche velaren Vokale unterster, unterer und mittlerer Zungenstellung sowie der palatale unterer und der palatale Labialvokal mittlerer Zungenstellung untersucht (langes [o:] kam insgesamt in zwei Fällen vor, sie wurden in der Analyse nicht berücksichtigt).

Für den Perzeptionstest wurden also die Vorkommen der genannten fünf Arten von Vokalen vorbereitet. Das abhörbereite Lautmaterial enthielt insgesamt 710 Vokale entsprechend der Verteilung in spontaner Rede, in Random-Reihenfolge (Tabelle 1). Der Test wurde in 50er Gruppen vorgenommen, mit Pausen. In einem Versuch wurden 200 Vokale abgehört.

Die Versuchspersonen waren Universitätsstudenten, ihr durchschnittliches Lebensalter 21 Jahre, insgesamt 10 Personen (6 Frauen, 4 Männer). Ihre Aufgabe war es, die Schriftzeichen der gehörten Vokale niederzuschreiben, es wurde also eine Methode gewählt, bei der die Entscheidung der Abhörer auf das ungarische Vokalsystem beschränkt blieb. Selbstverständlich war die Möglichkeit des Gebrauchs der Kategorie 'nicht identifizierbar' gegeben, die aber auf dem Testblatt keiner einzigen Versuchsperson vorkam. Die hinsichtlich ihrer reinen Phase segmentierten Vokale waren folglich im ungarischen Lautsystem eindeutig identifizierbar. Im Laufe der Aufarbeitung wurden die akustischen Angaben und die Perzeptionsangaben verglichen.

Tabelle 1: Zahl der aus spontaner Rede segmentierten Vokale

| Vokal (IPA-Umschrift) | Vorkommen | |
|--------------------------|-----------|---------|
| | Anzahl | Prozent |
| a: | 84 | 11,8 |
| 0 | 246 | 34,6 |
| o | 120 | 16,9 |
| ø, ø: | 39 | 5,4 |
| e | 221 | 31,3 |
| insgesamt | 710 | 100 |

Für die 710 Vokale wurden je vier akustische Parameter bestimmt sowie die Perzeptionsangabe der 10 Versuchspersonen, so daß insgesamt 28400 Daten für die Vokale entstanden. Bei den Analysen wurden die palatalen phonologisch kurzen und langen Vokale mittlerer Zungenstellung nicht voneinander getrennt. Als statistische Analysen wurden Signifikanzuntersuchung und Korrelationsberechnung vorgenommen.

ERGEBNISSE

Als Ausgangspunkt wurde bestimmt, welche die analysierten akustischen Angaben der Vokale [ø, ø:] in der Aussprache des untersuchten Sprechers sind (Tabelle 2). Den Testangaben gemäß haben alle Versuchspersonen diesen Sprechlaut zu 100% identifiziert. Die Durchschnittswerte für [ø] sind: Zeitdauer: 87 ms, F1: 410 Hz, F2: 1550 Hz, F3: 2320 Hz. Die Versuchspersonen haben ausnahmslos kurze [o]-Vokale identifiziert; zwischen den sprachlich kurzen und langen gibt es kaum einen physikalischen Zeitdauerunterschied. Der Durchschnitt der sprachlich kurzen beträgt 74,2 ms, der der langen 84,8 ms.

Tabelle 2: Zeitdauer- und Formantenwerte des Vokals [o]

| Zeitdauer (ms) | Formantenwerte (Hz) | | |
|----------------|---------------------|-----------|-----------|
| | F1 | F2 | F3 |
| 48-70 | 380-420 | 1440-1670 | 2000-2400 |
| 80-180 | 400-450 | 1350-1700 | 2260-2480 |
| 48-180 | 380-450 | 1350-1700 | 2000-2480 |

Im Laufe der weiteren Analysen wurden die Perzeptionsangaben als maßgeblich angesehen. Das Perzeptionsurteil der 10 Versuchspersonen galt dann für den gegebenen Vokal als gültig, wenn mindestens 8 Versuchspersonen identisch entschieden hatten (Minimum 80%). (Im Falle einer niedrigeren Perzeptionsidentität wurde der betreffende Vokal bei den Analysen nicht berücksichtigt.) Die Ergebnisse zeigen, daß die Versuchspersonen die untersuchten Vokale auf zweierlei Weise identifizierten: entweder erkannten sie den originalen Vokal oder sie vernahmen den kurzen Palatalvokal mittlerer Zungenstellung. Andere Lautqualitäten tauchten in den Wahrnehmungen nicht auf (im gesamten Korpus gab es nur im Falle von 5 Vokalen eine kleinere Perzeptionsidentität als 80%). In Tabelle (3) sind die Identifizierungsergebnisse für die 4 untersuchten Vokale aufgeführt. (Als Arbeitslösung wurden die auf den Perzeptionstestblättern mit *ö* bezeichneten Identifizierungen bei den Analysen als Schwa behandelt.)

Tabelle 3: Perzeptionsangaben der untersuchten Vokale

| Vokal | Identifizierung (%) | |
|------------------|---------------------|-----------|
| | als original | als Schwa |
| a: | 46,4 | 53,6 |
| o | 46,2 | 53,8 |
| o | 67,5 | 32,5 |
| e | 42,6 | 57,4 |
| durchschnittlich | 50,6 | 49,3 |

Die Daten der Tabelle zeigen, daß – aufgrund des gleichartigen Erkennens der Versuchspersonen – etwa die Hälfte der untersuchten Vokale in spontaner Rede als neutrale Lautqualität erscheinen. Zwischen [o] und den anderen drei Sprechlauten läßt sich ein signifikanter Unterschied feststellen, an letzteren gemessen bewahrt der Velarlaut mittlerer Zungenstellung am meisten die ursprüngliche Lautqualität. Die Ursache dafür liegt in der Artikulation: Unter Berücksichtigung dessen, daß der Velarvokal mittlerer Zungenstellung in nur einem Merkmal von dem Palatal mittlerer Zungenstellung abweicht, ist der Sprecher im Interesse der Unterscheidbarkeit wohl instinktiv um genauere Bildung bestrebt. Im Falle einer 2-Merkmal-Abweichung – zwischen [0] und [ø] hinsichtlich der horizontalen und senkrechten Bewegung der Zunge sowie zwischen [e] und [ø] hinsichtlich der senkrechten Bewegung der Zunge und der Lippenartikulation – ist die Artikulationsausführung wesentlich ungenauer, sie bemüht sich weniger um Unterschiedlichkeit. Die 3-Merkmal-Abweichung zwischen [a:] und [o] (die horizontale und senkrechte Bewegung sowie die Lippentätigkeit) erbringt ein ähnliches Ergebnis.

Im Laufe der Analysen wurden die Zeitdauer der Vokale und der Wert der ersten beiden Formanten mit den Perzeptionsangaben verglichen. Da der dritte Formant bei der Identifizierung der ungarischen Vokale eine redundante Rolle spielt, wurde bei den gegenwärtigen Analysen auf F3 nicht eingegangen. In Tabelle (4) werden die Streuungswerte für die akustische Struktur der Vokale angegeben.

Tabelle 4: Akustische Angaben der analysierten Vokale

| V | Zeitdauer (ms) | Formantengrenzwerte (Hz) | | Zeitdauer (ms) | Formantengrenzwerte (Hz) | |
|----|-------------------|-----------------------------|-----------|-------------------|-----------------------------|-----------|
| | | originaler Vokal | | | [«] | |
| | | F1 | F2 | | F1 | F2 |
| o | 30- 90 | 340-510 | 720-1100 | 40- 85 | 310-500 | 1075-1600 |
| 0 | 50-210 | 480-600 | 960-1230 | 25-155 | 390-540 | 1160-1500 |
| a: | 80-155 | 600-760 | 1335-1510 | 50-124 | 395-620 | 1415-1770 |
| e | 47-200 | 480-585 | 1460-1835 | 28-130 | 370-570 | 1060-1785 |

Die Angaben belegen zahlreiche relevante Unterschiede. Der Streubereich der Zeitdauer ist kleiner, die Werte sind niedriger beim Erkennen des neutralen Vokals, d.h., auch im Ungarischen ist bei der Veränderung der Lautqualität die kürzere Zeitdauer typisch. Bei Berücksichtigung sämtlicher als Schwa identifizierter Laute realisieren sich die ungarischen reduzierten Vokale zwischen ähnlichen Zeitdauer-Grenzwerten wie z.B. die holländischen. Für letztere sind Wertgrenzen von 18-120 ms bezeichnend, für die ungarischen solche von 25-130 ms; im Holländischen beträgt der häufige Zeitdauerwert 26-70 ms, mit einem

Durchschnittswert von 47 ms, im Ungarischen beträgt der Durchschnittswert 62 ms. Andererseits läßt sich nicht sagen, daß die Erkennung des Schwa nur eine Funktion des Zeitraumes sei, haben die Versuchspersonen doch auch Vokale kürzerer Zeitdauer entsprechend des ursprünglichen und auch solche mit längerer Zeitdauer als neutralen Vokal identifiziert (Abbildung 1).

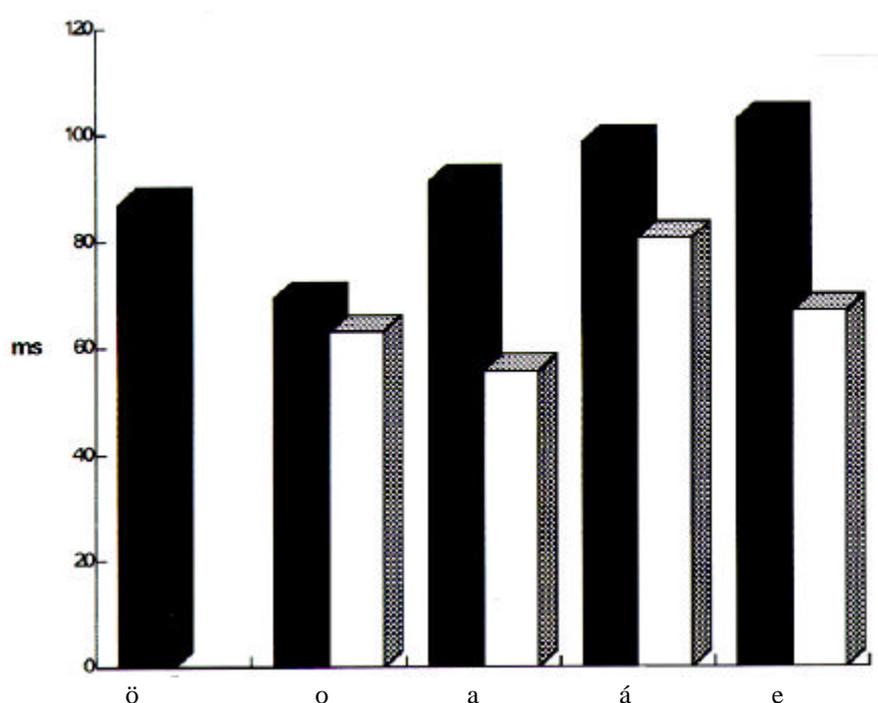


Abbildung 1: Die Gestaltung der Zeitdauerdurchschnitte im Falle zweifacher Identifizierung (schwarz-originale Vokale, weiss-Schwa)

Die Analyse der Formantenwerte bot weitere Informationen zur Erklärung der Identifizierungen. Einen charakteristischen Unterschied zeigen die Grenzwerte des ersten Formanten betreffs des originalen und des neutralen Vokals. Der Frequenzbereich der als Schwa identifizierten Laute ist wesentlich enger und realisiert sich niedriger. Vergleicht man den F1-Wert sämtlicher als neutral identifizierter Vokale und die Angaben über die ersten Formanten der [ø]-Aussprache desselben Sprechers ($\emptyset F1 = 380-450$ Hz), ergibt sich, daß mit Ausnahme der Schwa-Wahrnehmung des originalen [a:]-Vokals die Grenzwerte sehr ähnlich sind: $oF1 = 310-500$ Hz, $0F1 = 390-540$ Hz und $eF1 = 370-570$ Hz. Die Werte der zweiten Formanten beim originalen [ø]-Vokal: $F2 = 1350-1675$ Hz, demgegenüber ist dieser Bereich bei den als Schwa identifizierten Lauten wesentlich breiter, bei der Verschiebung sowohl in Richtung niedrigerer als auch höherer Frequenzen. Das weist eindeutig darauf hin, daß die Artikulationseigenarten der original ausgesprochenen Vokale sich in gewissem Maße im

Wert der zweiten Formanten bewahren. Die für den betreffenden Sprecher typische Akustikstruktur des [o] erscheint dagegen in geringerem oder größerem Maße in jedem als Schwa identifizierten Vokal. Die im untersuchten Korpus für Schwa bekommenen F1-Grenzwerte sind 313-622 Hz, der Durchschnitt ist 486 Hz, die F2-Grenzwerte sind 886-1785 Hz und der Durchschnitt 1457 Hz. Für die Formantenwerte, die für das nicht als Phonem-Realisation erscheinende Schwa – mit traditioneller Methode – gemessen worden sind, wurden folgende Angaben mitgeteilt: F1 = 210-660 Hz, F2 = 1050-2410 Hz (vgl. Kassai 1978: 95). Zum Vergleich: Der F1-Durchschnitt der holländischen neutralen Vokale ist 335 Hz, der F2-Durchschnitt 1285 Hz (Koopmans-van Beinum op. cit.); die der englischen: F1 = 480-580 Hz, F2 = 1350-1530 Hz (Fry 1976: 93). In unserem Material unterscheiden sich die F1- und F2-Werte der als original und der als neutral erkannten Vokale signifikant (Abbildung 2 und 3).

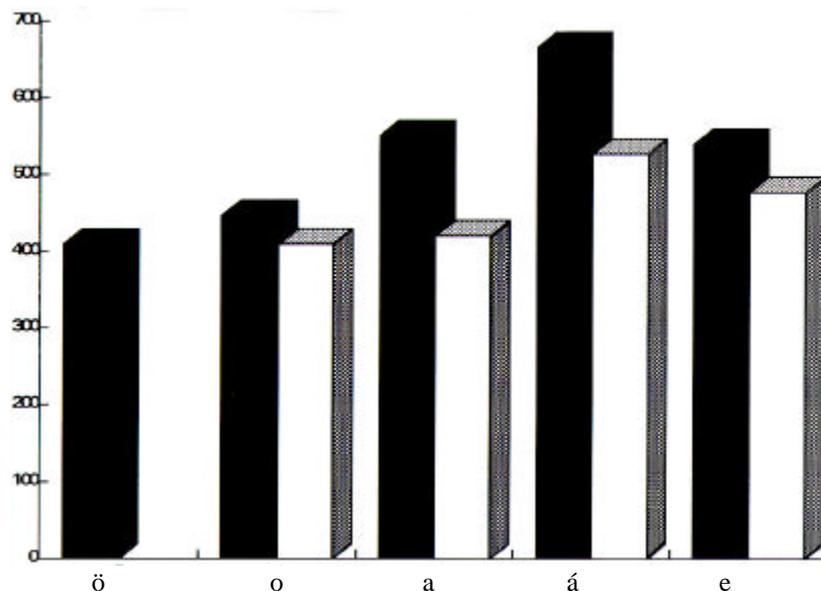


Abbildung 2: Die Durchschnittswerte des ersten Formanten in Funktion der Identifizierung (schwarz-originale Vokale, weiss-Schwa)

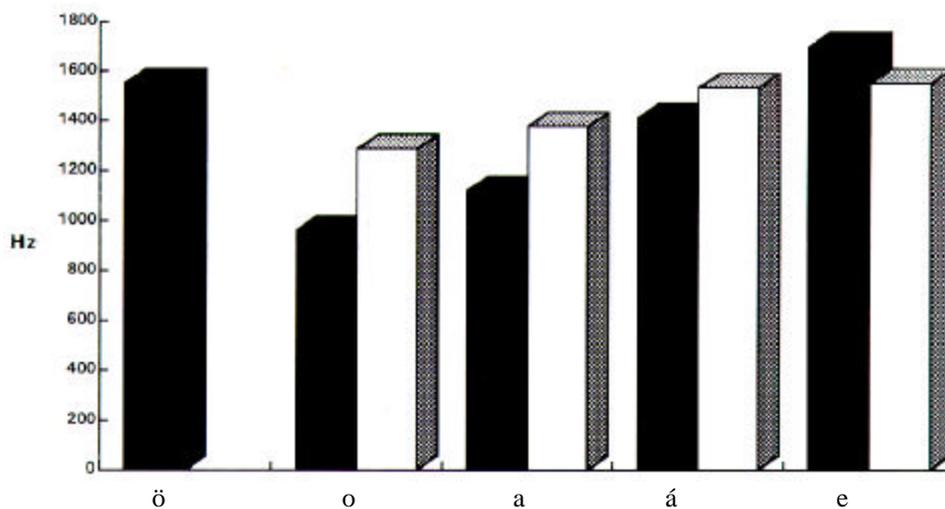


Abbildung 3: Die Durchschnittswerte des zweiten Formanten in Funktion der Identifizierung (schwarz-originale Vokale, weiss-Schwa)

Mittels mathematischer Korrelationsberechnung wurde versucht, eine Antwort zu bekommen, wie sich die drei analysierten Veränderlichen (Zeitraum, F1 und F2) zur Perzeption des gegebenen Vokals verhalten. Bei welchem Faktor findet sich ein enger und bei welchem ein weniger enger Zusammenhang, bzw. ist eventuell der betreffende Faktor nicht determinierend? In Tabelle (5) sind die Ergebnisse der Korrelationsberechnung zusammengefaßt, die Werte sind bis zur 5. Dezimalstelle angegeben (Abkürzungen: Zeit = physikalische Zeitdauer, Perzep. = Perzeption, Wahrnehmung, r = Korrelationswert, Korr. = Korrelation).

Tabelle 5: Die Zusammenhänge von akustischen Angaben und Perzeption

| Vokal | Korrelation | r | Interpretation |
|-------|--------------|----------|----------------|
| o | Zeit/Perzep. | 0,17891 | keine Korr. |
| | F1/Perzep. | 0,31886 | mittlere Korr. |
| | F2/Perzep. | -0,74000 | starke Korr. |
| | Zeit/F1 | 0,22106 | keine Korr. |
| | Zeit/F2 | -0,20817 | keine Korr. |
| 0 | Zeit/Perzep. | -0,38644 | mittlere Korr. |
| | F1/Perzep. | -0,52427 | starke Korr. |
| | F2/Perzep. | 0,40336 | mittlere Korr. |
| | Zeit/F1 | 0,37416 | mittlere Korr. |
| | Zeit/F2 | -0,2376 | keine Korr. |
| a: | Zeit/Perzep. | 0,43666 | mittlere Korr. |
| | F1/Perzep. | 0,76717 | starke Korr. |

| | | | |
|---|--------------|----------|----------------|
| | F2/Perzep. | -0,63991 | starke Korr. |
| | Zeit/F1 | 0,43601 | mittlere Korr. |
| | Zeit/F2 | -0,40085 | mittlere Korr. |
| e | Zeit/Perzep. | -0,42007 | mittlere Korr. |
| | F1/Perzep. | -0,50174 | starke Korr. |
| | F2/Perzep. | -0,42817 | mittlere Korr. |
| | Zeit/F1 | 0,315221 | mittlere Korr. |
| | Zeit/F2 | 0,294632 | keine Korr. |

Die Identifizierung von Zeitdauer und Vokalqualität zeigt mit Ausnahme des [o] mittlere Korrelation, d.h., die Zeitdauer hat Einfluß auf die Perzeption. Dabei korreliert im Falle von [o] die Zeitdauer nicht mit der Perzeption. Der Wert des ersten Formanten und die wahrgenommene Lautqualität zeigen wiederum mit Ausnahme des [o] – bei dem die Korrelation mittelmäßig ist – im Falle der drei anderen Vokale starken Zusammenhang. Der Wert des zweiten Formanten und die Wahrnehmung zeigen starke Korrelation im Falle von [o] und [a:], mittlere Korrelation im Falle von [0] und [e]. Insgesamt läßt sich also sagen, daß die Wahrnehmung die Qualität des originalen Vokals definiert, wobei aber in größtem Maße der Wert der ersten Formanten, dann der F2 und schließlich die Zeitdauer eine Rolle spielen.

Obwohl das Vorkommen betonter Vokale niedrig ist, wurde untersucht, ob die erhaltenen Angaben eine erhebliche Abweichung in Funktion des Erscheinens der Betonung zeigen. Das ganze Material als 100% genommen, beträgt der Anteil sämtlicher betonter Vokale 24,2%; diesen Anteil als 100% genommen, fallen auf [o] 30,9%, auf [0] 15,6%, auf [a:] 26% und auf [e] 27,5%. Die Anteile ändern sich unter Berücksichtigung des Anteils der betonten bei den unterschiedlich wahrgenommenen Vokalen (Tabelle 6).

Tabelle 6: Identifizierung der betonten und unbetonten Vokale

| Vokal (original/wahrgenommen) | Vorkommen (%) | |
|----------------------------------|---------------|----------|
| | betont | unbetont |
| o/o | 28,5 | 71,5 |
| o/« | 33,3 | 66,6 |
| 0/0 | 16,6 | 83,4 |
| 0/« | 10,9 | 89,1 |
| a:/a: | 50 | 50 |
| a:/« | – | 100 |
| e/e | 42,8 | 57,2 |
| e/« | 12,2 | 87,8 |

Die betonten Vokale erkennen die Versuchspersonen – innerhalb der Relationen – in größerem Maße bei den originalen Vokalen, was besonders typisch für [a:] und [e] ist. Kein signifikanter Unterschied besteht bei [a] und [o] (bei letzterem sind die als Schwa identifizierten Laute mit einem etwas größeren Anteil in betonter Stellung). Eigenartig gestalten sich die Zeitdauerwerte in Funktion der betonten Stellung (Tabelle 7).

Tabelle 7: Die physikalische Zeitdauer bei betonten/unbetonten Vokalen

| Vokal (original/wahrgenommen) | Physikalische Zeitdauer (ms) | |
|----------------------------------|------------------------------|----------|
| | betont | unbetont |
| o/o | 56 | 71,6 |
| o/« | 72,5 | 61 |
| 0/0 | 85,5 | 94,2 |
| 0/« | 87,8 | 66,6 |
| a:/a: | 122 | 87,6 |
| a:/« | – | 87,5 |
| e/e | 104,2 | 101,2 |
| e/« | 71,2 | 66,3 |

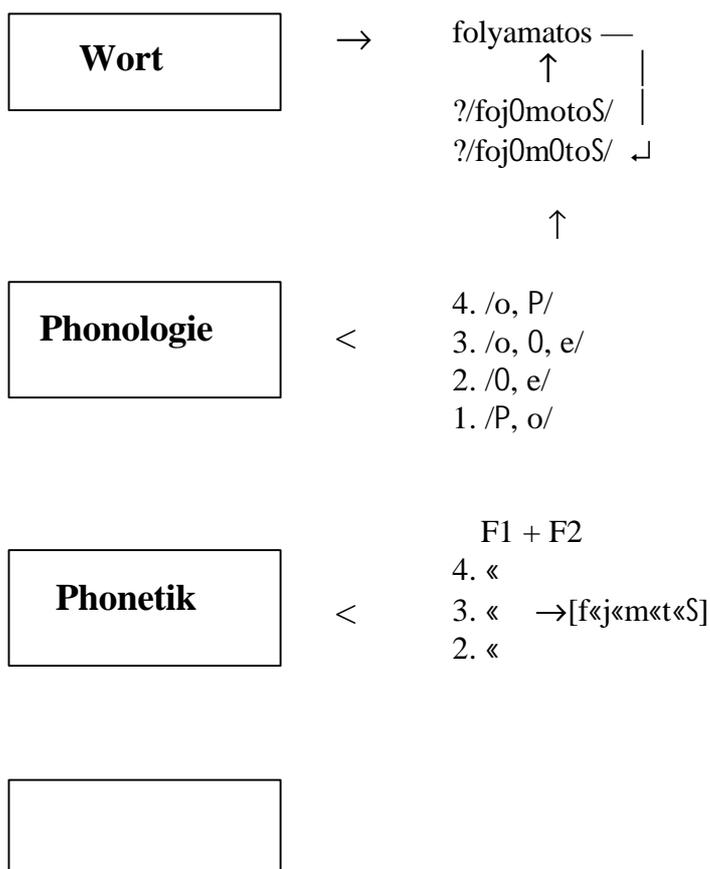
Die Hypothese lautete, daß die physikalische Zeitdauer der betonten Vokale die der unbetonten erheblich übersteigen wird, obwohl dies von der Gestaltung des [0] und [e] in absoluter Wortendstellung beeinflusst werden kann (vgl. Kassai 1982). Im Falle der velaren Laute mittlerer und unterer Zungenstellung hat sich die Hypothese nicht völlig bestätigt: Bei Bewahrung der originalen Lautqualität war die Zeitdauer in unbetonter Stellung größer als in betonter; gleichzeitig verringerte sich aber bei Identifizierung des Schwa die Zeitdauer der unbetonten erheblich. Im Falle von [a:] waren die in unbetonter Stellung kürzer (zu bemerken ist, daß hier die Angaben beschränkt waren). Im Falle von [e] zeigt sich kaum ein Unterschied in der Zeitdauer bei Identifizierung des originalen Vokals, aber auch im Falle der Identifizierung des Schwa ist der Unterschied in der Zeitdauer nicht signifikant.

Die andere Analysemöglichkeit zeigt, daß die Zeitdauerwert-Durchschnitte der Laute in unbetonter Stellung – mit Ausnahme von [a:] – im Falle der Identifizierung des neutralen Vokals kleiner sind. Bei denen in betonter Stellung gab es keinen Unterschied bei [0] und ein umgekehrtes Verhältnis bei [o]. Die Formantenwerte zeigten im Falle keines einzigen Vokals im Material eine signifikante Abweichung abhängig von betonter Stellung (was sich auch aus der relativ kleinen Zahl der betonten Laute ergeben kann).

Untersucht wurde, in welchem Verhältnis die vom Gewährsmann ausgesprochenen Vokale [ø, ø:] hinsichtlich des Bereiches F1/F2 zum Bereich F1/F2 der als Schwa identifizierten Vokale stehen. Die Ergebnisse zeigen, daß die F1/F2-Struktur der als Schwa

identifizierten Vokale mehr oder weniger mit dem akustischen Bereich der originalen Palatalvokale mittlerer Zungenstellung übereinstimmt, aber charakteristischerweise zugleich auch die für die originalen Vokale typische Struktur bewahrt. Sowohl die Werte des ersten als auch des zweiten Formanten enthalten den für die originale Lautqualität typischen Wert. Während z.B. das F1 des in der Stellung des originalen [0] erscheinenden Schwa bei etwa 500 Hz erscheint (bei korrekter Identifizierung sind die Grenzwerte des ersten Formanten des Lautes 483-600 Hz), verschiebt sich der Durchschnittswert des anstelle von [a:] erscheinenden Schwa zu 600 Hz hin (Grenzwerte bei korrekter Identifizierung 600-760 Hz.). Der F2-Wert des Schwa anstelle von [e] verschiebt sich bis auf 1800 Hz, dagegen der des neutralen Vokals anstelle von [0] nur etwa auf 1600 Hz.

Abbildung (4) stellt – in Kenntnis all dessen – ein mögliches Modell der Worterkennung im Falle spontaner Rede dar. Gezeigt wird der Prozeß der Aufarbeitung des Wortes *folyamatos* im Laufe der Worterkennung. Auf akustischer Ebene wurden – hier nur der Einfachheit halber – die Frequenzwerte der ersten beiden Formanten aufgeführt. Die phonetische Ebene identifiziert aufgrund der akustischen Analyseergebnisse die Sprechlaute, im Falle sämtlicher Vokale das neutrale Schwa. Die Phonemebene zählt aufgrund aller bisheriger Ergebnisse die möglichen "originalen" Vokalphoneme auf und wählt mittels Anwendung der ungarischen phonotaktischen Regeln die einzig möglichen aus. Die Entscheidungsprozesse der Wortebene bestimmen dann aufgrund der im mentalen Lexikon gespeicherten Einheiten endgültig das gegebene Wort.



| | | | | |
|----------------|---|------|--------|---------|
| | | 1. « | | |
| | | | F1 | F2 (Hz) |
| | | | 4. 370 | 1420 |
| | | | 3. 490 | 1440 |
| Akustik | < | | 2. 530 | 1370 |
| | | | 1. 350 | 1300 |

Abbildung 4: Das Wort *folyamatos* im Prozeß der Worterkennung

FOLGERUNGEN

Die in den Wörtern des spontanen Redematerials des Gewährsmannes vorkommenden analysierten Vokale haben die Hypothese zweifelsfrei bewiesen: Ein großer Teil der vom Sprecher/Hörer als vollwertige Vokale wahrgenommenen Sprechlaute ist tatsächlich nicht mit dem vermuteten Vokal identisch. Es war vorausgesetzt worden, daß ein Grund für die Veränderung in Richtung der neutralen Lautqualität die Beschleunigung des Sprechtempo sei. Dies hat die Instrumentenanalyse nur zum Teil untermauert, da das Sprechtempo des Gewährsmannes zwischen einem langsamen und dem durchschnittlichen Artikulationstempo wechselte; die Umschläge der Lautqualität realisierten sich dennoch in erheblichem Maße. Das Sprechtempo selbst konnte die Veränderung also nur geringfügig verursachen. Die zweite Hypothese hinsichtlich des Grundes war die Tatsache der spontanen Rede selbst. Anscheinend weist das nahezu 50%ige Auftreten der erhaltenen Angabenmenge, der neutralen Vokale, in der spontanen Rede darauf hin, daß der/die tatsächliche/n Grund/Gründe in der fortlaufenden Artikulation bzw. in der Spontaneität der Sprechplanung zu suchen sind.

Im Falle spontaner Rede ist die Gestaltung der Gedankeninhalte in diskreten sprachlichen Einheiten (Wörtern, Wortkonstruktionen usw.) im Redeplanungsprozeß bestimmend. Das hat zur Folge, daß für die Planung der Artikulationsbewegungsreihen wesentlich weniger Zeit und Aufmerksamkeit bleibt, die hierbei ablaufenden Transformationen sind schneller als beispielsweise beim lauten Lesen (wo sämtliche Transformationen in jeder Hinsicht viel ausgeglichener sind). Im Falle der vorher nicht geplanten Kommunikation steht die Gestaltung des gedanklichen Inhaltes in diskreten sprachlichen Einheiten im Mittelpunkt, weshalb sich die Artikulationsbewegungsreihen als sehr automatische Operationsserie realisieren. Anders formuliert, ist das erstrangige Ziel des Sprechers die sprachliche Gestaltung des gedanklichen Inhaltes, und erst danach, also sekundär, erscheint die Aufgabe der aktuellen Hörbarmachung. Die primäre Transformationsserie ist jederzeit individuell, handelt es sich doch in den meisten Fällen um

eine Generierung früher so noch nicht formulierter sprachlicher Konstruktionen. Die Hörbarmachung generierter sprachlicher Konstruktionen ist jedoch eine – zumindest auf der segmentalen Ebene der Artikulation – wiederholte Kette eingeübter Gesten. Mit einem Beispiel aus dem Material des Gewährsmannes bedeutet dies aufgrund des analysierten Materials folgendes. Ein Abschnitt aus dem Material des Gewährsmannes, ein Text zwischen zwei längeren "Denk-", d.h. Sprechplanungspausen (die kürzeren Sprechpausen gliedern die Mitteilung auch syntaktisch):

"... vannak köztük kaland vannak akciók ... általában különböző típusú gyerekeknek készült ... vannak olyanok amik inkább feszültségvezető folyamatos izgalmat biztosítanak ..." (... es gibt darunter Abenteuer es gibt Aktionen ... im allgemeinen für Kinder verschiedenen Typs verfertigt ... es gibt solche die eher spannungsableitend sind ständige Aufregung garantieren ...)

Die syntaktische Verwirklichung des Textabschnittes, die Schaffung des gegebenen Inhaltes mittels vorliegender/n Konstruktion/en wurde individuell, unter den gegebenen Umständen einer gegebenen Situation formuliert. Doch sind die lexikalischen Einheiten, die Syntagmen der Struktur auf der Artikulationsebene mit großer Wahrscheinlichkeit sehr oft geschaffene Operationsserien. Es ist nicht anzunehmen, daß der Gewährsmann diesen Satz schon irgendwann ausgesprochen hat (aber in derselben sprachlichen Form zumindest nicht oft), wie es – andererseits – nicht anzunehmen ist, daß er diese die Mitteilung bildenden Wörter, wie *vannak*, *készült*, *olyanok*, *izgalmat* oder *gyerekek*, nicht schon (oftmals) ausgesprochen hätte. Die sprachliche Transformierung des Gedankens ist also keine Serie von Klischees, die Aussprache aber wohl, selbst dann, wenn die Anpassung (die Koartikulation der nebeneinander geratenen Wörter) seltener vorkommt, eventuell so zum ersten Mal.

Nach Angaben von 1981 treten in der deutschen Sprache die neutralen Vokale mit 23,88%, in der englischen mit 22,99% und in der französischen mit 7,6% auf (Onishi 1981). Diese Werte spiegeln die regulären (ihrem Phonem entsprechenden) Schwa-Realisationen der Rede wider. Es gibt keine Angaben darüber, wie sich in diesen Sprachen im Falle spontaner Rede die Lautqualität der übrigen Vokale verändert, in welchem Anteil auch diese eventuell in den neutralen Vokal umschlagen.

Aufgrund der vorliegenden Versuchsergebnisse scheint die Revidierung der – heute schon als klassisch geltenden – Feststellung über die Qualität der ungarischen Vokale berechtigt zu sein: In spontaner Rede ändert sich die Qualität eines Teils von ihnen, sie enthalten die akustischen Eigenheiten des neutralen Schwa, und das weist eindeutig auf den klischeehaften Abrufprozeß der spontanen Redeauführung hin. Die Erscheinung tritt nicht nur im Falle der unbetonten, sondern auch der betonten Vokale auf, und das hängt nur

teilweise von der aktuellen (physikalischen) Zeitdauer ab. Aus der Sicht der Sprachrichtigkeit und der Aussprache-Stilistik wurden hier vermutlich die Ergebnisse einer unerwünschten Veränderung registriert.

LITERATUR

- Bolla, K. 1982 Magyar hangalbum (Ungarisches Lautalbum), in: Bolla, K. (Hg.) *Fejezetek a leíró magyar hangtanból*, Budapest: 165-173.
- Fry, D. (ed.) 1976 *Acoustic Phonetics*, Cambridge.
- Gósy, M. 1992 *Speech Perception. Forum Phonetikum*, Frankfurt/M.
- Jones, D. 1966 *The Pronunciation of English*, Cambridge.
- Kassai, I. 1978 Sva-jelenségek a magyar beszédben (Schwa-Erscheinungen in der ungarischen Rede), *Magyar Fonetikai Füzetek* 1: 92-96.
- 1982 Időtartam és kvantitás a magyarban (Zeitdauer und Quantität im Ungarischen), in: Bolla, K. (Hg.) *Fejezetek a leíró magyar hangtanból*, Budapest: 115-145.
- Koopmans-van Beinum, F.J. 1994 What's in a schwa? Durational and spectral analysis of natural continuous speech and diphones in Dutch, *Phonetica* 51: 68-80.
- Laziczius, G. 1944/1963 *Fonetika* (Phonetik), Budapest.
- Lindner, G. 1981 *Grundlagen und Anwendung der Phonetik*, Berlin.
- Magdics, K. 1965 *A magyar beszédhangok akusztikai szerkezete* (Die akustische Struktur der ungarischen Sprechlaute), Budapest.
- Molnár, J. 1970 *A magyar beszédhangok atlasza* (Atlas der ungarischen Sprechlaute), Budapest.
- Olaszy, G. 1985 *A magyar beszéd leggyakoribb hangsorépítő elemeinek szerkezete és szintézise* (Struktur und Synthese der häufigsten lautreihenbauenden Elemente der ungarischen Rede), Budapest.
- Onishi, M. (ed.) 1981 *A Grand Dictionary of Phonetics*, Tokio.
- Szalai, E. 1995 Az [u:], [a:] és [i] hangok koartikulációs mezőiről (Über die Koartikulationsfelder der Laute [u:], [a:] und [i]), in: *Beszédkutatás* 95, hg. v. M. Gósy, Budapest: 83-93.
- Tarnóczy, T. 1941 *A magyar magánhangzók akusztikai szerkezete* (Die akustische Struktur der ungarischen Vokale), Budapest.
- 1965 Acoustic analysis of Hungarian vowels, *QPSR Januar 1965*, Stockholm: 8-12.

Mária Gósy
UAW Institut für Sprachwissenschaft, Budapest