

# Emotionalität, stimmlicher Ausdruck und Sprechweise

Sylvia Moosmüller, Ralf Vollmann & Werner A. Deutsch

## 1. Daten

Als wegweisend für die Analyse stimmlichen Ausdrucks in Österreich sind wohl die Arbeiten Felix Trojans zu nennen. Aus seinem umfassenden Datenmaterial zur Gefühlsresonanz, das er sowohl mit Schauspielerinnen und Schauspielern als auch mit seiner eigenen Stimme gesammelt hat, und das am Phonogrammarchiv der ÖAW archiviert vorliegt, haben wir bislang folgende Emotionen einer phonetischen und phonologischen Analyse unterzogen (wir verwenden hier Trojans Terminologie):

↳ Verstandesmäßiger Ausdruck (VA):

„Eine Aussage, daß die Größen a und b einander gleich sind, nennen wir eine Gleichung und a und b ihre Seiten oder Teile.“

↳ Vereinigungsstreben (VS):

„Bist du aber lieb. Komm, gib mir deine Hand und laß dir einen Kuß geben.“

↳ Aggressive Ergotropie (AE):

„Noch ein Wort und ich hau Ihnen eine Ohrfeige herunter, Sie unverschämter Mensch Sie! So eine Frechheit ist mir doch in meinem Leben noch nicht untergekommen!“

Verglichen wurde dieses simulierte Material mit authentischem Material<sup>1</sup>, nämlich mit Sätzen aus verschiedenen Telefongesprächen, die von der angerufenen Person aufgezeichnet wurden. Vom emotionalen Zustand her sind diese Daten den Daten Trojans vergleichbar.

↳ Verstandesmäßiger Ausdruck (VA):

„Des wa:ß jo i: net, wos de mitanand khort ham vorher. Na, he, huach amoe, i: muaß den Hawerer irgendwo derglenga<sup>2</sup> morgen.“

↳ Vereinigungsstreben (VS):

„Weil, des howi da eh scho ksogt, daß du mein Herzblatt bist. Na, dann sag i: da´s jetzt. Verstehst, i: ma:n, du bist a:foch die Frau!“

↳ Aggressive Ergotropie (AE):

„I: ruf seit aner halben Stund bei dir an, he, i: ma:n, bist deppert oder wos?“

---

<sup>1</sup> Es handelt sich um einen Sprecher des österreichischen Dialekts im Westen Oberösterreichs (mittelbairisch).

<sup>2</sup> erreichen.

## 2. Hochsprache und Dialekt

Wir gehen davon aus, daß sowohl die Hochsprache als auch der Dialekt funktional gleichberechtigte Systeme sind. Dialekt und Hochsprache bewegen sich demnach nicht auf einer graduellen Skala von unten nach oben, in der der Dialekt die informelle, „schlampige“ Variante darstellt und die Hochsprache die formelle, artikulationsaufwendigere Form, sondern es handelt sich um zwei gleichberechtigte Systeme, die erlernt werden müssen, wobei jeweils das nicht in der Primärsozialisation erlernte System das für den Sprecher / die Sprecherin artikulationsaufwendigere ist.

Insofern ist es auch unrichtig, von funktional und situationell unterschiedlichen Systemen zu sprechen, denn die funktionelle Trennung ergibt sich erst sekundär durch die unterschiedlichen Funktionen, die die jeweiligen Sprecher/innen des einen oder anderen Systems meistens innehaben, sowie aus den jeweiligen Situationen, in denen sie sich meistens bewegen.

Es ergibt sich daraus zwangsläufig, daß der Dialekt nicht primär für den emotionalen und subjektiven, und die Hochsprache für den abstrakten, objektiven Sprachgebrauch zuständig ist, sondern in beiden Systemen finden die Sprecher/innen die Möglichkeit, sowohl Subjektives als auch Objektives auszudrücken oder formelle und informelle Situationen zu meistern. Genauso, wie also beide Systeme einen formellen und einen informellen Bereich haben, haben auch beide Systeme dieselben Möglichkeiten, verschiedene emotionale Zustände zum Ausdruck zu bringen (vgl. Moosmüller 1991, Moosmüller/Vollmann 1994).

Es verfügen somit beide Systeme über – teilweise dieselben – Möglichkeiten von Entdeutlichungs- oder Verdeutlichungsprozessen<sup>3</sup>, beide Systeme sind gekennzeichnet durch Reduktions-, Assimilations- und Tilgungsprozesse in unbetonten Positionen oder eben durch Verdeutlichungsprozesse in betonten Positionen. Insofern kann ein hochsprachliches /ha:bEn/ und ein dialektales /hO:m1/ bis zu einem silbischen Nasal [m1] reduziert werden.

Aus diesem Grund ist es möglich, die Gemeinsamkeiten und Unterschiede von simuliertem hochsprachlichen und authentischem dialektalen Material zu untersuchen.

## 3. Ergebnisse

Die untersuchten Signale wurden mit 16 kHz Samplingrate und 16 Bit mono linear digitalisiert und auf Computerplatte gespeichert. Die Segmentierung und die instrumentalphonetische Analyse wurden an einer S\_Tools Work Station (vgl. Deutsch & Noll 1990, 1994) durchgeführt; die Formantfrequenzkandidaten wurden mittels der LPC-Analyse (konstant mit 20 Koeffizienten und einer Preemphasis von 0,9, frame length 64 ms, frame shift 8 ms) ermittelt, die Grundfrequenz wurde nach der Methode SIFT (vgl. Markel & Grey 1976) extrahiert.

### 3.1. Durchschnittliche Lage der Sprachgrundfrequenz

Die größten Unterschiede zwischen authentischem und simuliertem Material sind in der durchschnittlichen Lage der Stimmgrundfrequenz zu beobachten (vgl. Tab. 1):

	simuliert	authentisch
VA	125 Hz	147 Hz
AE	160 Hz	230 Hz
VS	99 Hz	130 Hz

Tab. 1: *Mittelwerte der Sprachgrundfrequenz bei simuliertem und authentischem Material.*

Wie aus Tabelle 1 ersichtlich, ist der Anstieg der durchschnittlichen Lage der Sprachgrundfrequenz bei AE gegenüber VA beim authentischen Material wesentlich höher als beim simulierten Material. Der Unterschied zwischen AE und VA beträgt beim simulierten Material 427 cents, beim authentischen Material hingegen 770 cents! Diese Ergebnisse sind nicht deckungsgleich mit den Ergebnissen von Williams & Stevens (1972), bei ihrer

<sup>3</sup> Zur Theorie der Natürlichen Phonologie vgl. Dressler 1984

Untersuchung betrug der Unterschied beim authentischen Material lediglich 287 cents, beim simulierten Material hingegen 825 cents<sup>4</sup>.

Auch bei VS sind die Unterschiede zwischen authentischem und simuliertem Material beträchtlich; die Differenz gegenüber VA beträgt beim authentischen Material 400 cents, beim simulierten Material hingegen nur 212 cents.

### 3.2. Umfang und Dynamik der Sprachgrundfrequenz

Beim simulierten Material weist AE gegenüber VA und VS perzeptuell den größten Umfang in der Sprachgrundfrequenz auf. Beim authentischen Material hingegen weist VA den größten Umfang auf, AE und VS zeigen einen perzeptuell geringeren Umfang (vgl. Tabelle 2):

	simuliert (in cents)	authentisch (in cents)
VA	1885	1720
AE	2343	1357
VS	1464	1102

Tab. 2: *Bereich der Sprachgrundfrequenz zwischen dem höchsten und den tiefsten gemessenen Wert bei simuliertem und authentischem Material (in cents).*

Der geringere Umfang in der Sprachgrundfrequenz bei AE (authentisches Material) erklärt sich natürlich aus der durchschnittlich höheren Lage. Die Unterschiede zwischen authentischem und simuliertem Material gerade bei AE sind aber beachtlich (siehe Diagramm 1): Der Sprecher des simulierten Materials scheint die bereits festgestellten Unterschiede bei der durchschnittlichen Lage durch einen perzeptuell wesentlich größeren Umfang auszugleichen, der sich selbstverständlich auch auf die Dynamik auswirkt. So beträgt der maximal gemessene Unterschied zwischen betonter und unbetonter Silbe 1558 cents in „ich hau“ (simuliertes Material), hingegen nur 1046 cents in „an“ (authentisches Material). Ähnlich hohe Unterschiede sind bei AE des simulierten Materials in jedem Wechsel von betonter zu unbetonter Silbe zu beobachten, also bei „noch ein“ (1156 cents), „herunter“ (655 cents), „Sie unverschämter“ (1293 cents), „eine Frechheit“ (1284 cents), „meinem Leben“ (781 cents). Diese Unterschiede sind beim authentischen Material wesentlich geringer, nämlich 595 cents bei „aner halben“, 619 cents bei „Stund“, 149 cents bei „he i man“, 1008 cents bei „bist deppert“ und 824 cents bei „was“, es kommt also nur einmal zu einer Differenz von 10 Halbtönen, eine Differenz von einer Oktave ist nie beobachtbar.

Interessant ist noch eine relative Beruhigung gegen Ende der Äußerungseinheit beim simulierten Material. Hier sinkt die Sprachgrundfrequenz kontinuierlich ab. Die relative Beruhigung am Ende der Äußerungseinheit der simulierten AE ist bei diesem Beispiel nicht vergleichbar, da das Beispiel der authentischen AE in einer Frage endet, wir möchten aber darauf hinweisen, daß dies auch bei authentischem Material zu beobachten ist (siehe Diagramm 1):

<sup>4</sup> Williams & Stevens (1972) untersuchten die Beschreibung des Absturzes der Hindenburg, und zwar vor und nach dem Absturz. Die Daten sind mit den unsrigen insofern nicht vollkommen vergleichbar, da es sich um die Beschreibung einer Katastrophe handelt und nicht um den emotionalen Zustand „Ärger“. Ferner muß in Rechnung gestellt werden, daß gerade Nachrichtenreporter in der Beherrschung von Emotionen geschult werden. Die absoluten Werte (Median) sind Williams & Stevens (1972: 1246) entnommen. Nach Bezoooyen (1984) korrelieren bei der Messung der Sprachgrundfrequenz Median und Mittelwert hoch, sodaß beide Methoden in gleicher Weise aussagekräftig sind.

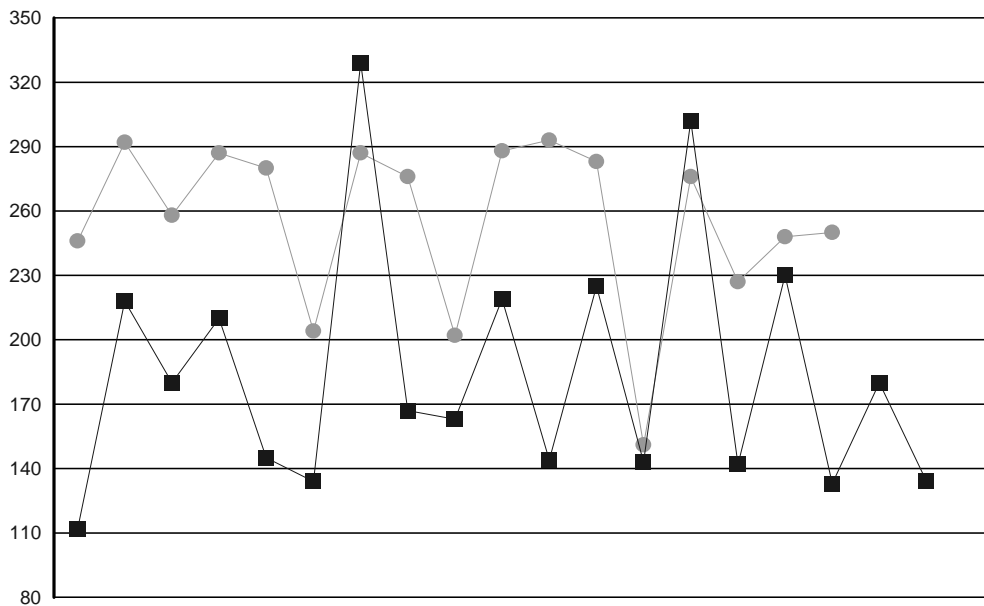


Diagramm 1: Sprachgrundfrequenzverlauf bei simuliertem (—■) und authentischem Material (—●). y-Achse: Frequenz in Hz.

Auch bei VS ist der Umfang in der Sprachgrundfrequenz beim authentischen Material geringer als beim simulierten. Interessant ist in beiden Beispielen der Anstieg der Sprachgrundfrequenz am Ende der Äußerungseinheit (vgl. dazu Schmidt's Ansatz der hörerrelevanten funktionalen Kategorien (Schmidt 1994)).

Lediglich bei VA scheinen die Unterschiede vernachlässigbar, obwohl beim simulierten Material wieder die perzeptuelle Differenz zwischen betonten und unbetonten Silben größer sind als beim simulierten Material: 725 cents bei „eine Aussage“, 667 cents bei „Größen“, 570 cents bei „einander gleich“, 759 cents bei „und a“, 835 cents bei „ihre Seiten“ vs. 494 cents bei „waß ja“, 149 cents bei „was de“, 204 cents bei „khoft habm“, 196 cents bei „he horch“, 532 cents bei „i muaß“, 257 cents bei „Hawerer“. In beiden Fällen ist ein Abfall der Sprachgrundfrequenz am Ende der Äußerungseinheit zu beobachten.

### 3.2. Gipfelmuster

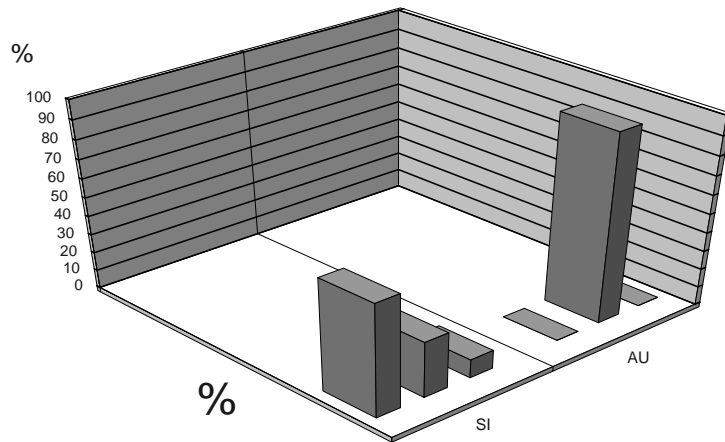
Nach Gartenberg & Panzlaff-Reuter (1991) können frühe, mittlere und späte Gipfelmuster unterschieden werden, die jeweils auch perzipierbar sind (Kohler & Gartenberg 1991). Perzeptionstests zeigten, daß mittlere Gipfel gegenüber frühen und späten Gipfeln als betonter wahrgenommen werden, und zwar aufgrund der höheren Energiekonzentration in der Mitte des Signals. Aus diesem Grund ist bei mittleren Gipfeln auch ein geringerer Unterschied zwischen Anstieg, Gipfel und Abfall notwendig, um den Eindruck der Betontheit zu erwecken, als dies bei frühen oder späten Gipfeln der Fall ist.

Beim simulierten Material ist auffällig, daß die betonten Silben bei AE überwiegend einen mittleren Gipfel aufweisen, die betonten Silben bei VS überwiegend einen frühen und die betonten Silben bei VA überwiegend einen späten Gipfel. Bei AE wird also der Eindruck der Betontheit durch die Verwendung mittlerer Gipfel verstärkt, die zusätzlich noch einen großen Unterschied zwischen Anstieg, Gipfel und Abfall aufweisen.

Beim authentischen Material sind diese Unterschiede zwischen den drei emotionalen Zuständen weit weniger auffällig. Bei AE realisiert der Sprecher nur späte Gipfel, bei VS überwiegend späte Gipfel (zu 50%) und bei VA überwiegend frühe Gipfel (vgl. Diagramm 2a-2c):

### GIPFELMUSTER (AE)

*mittel - spät - früh*

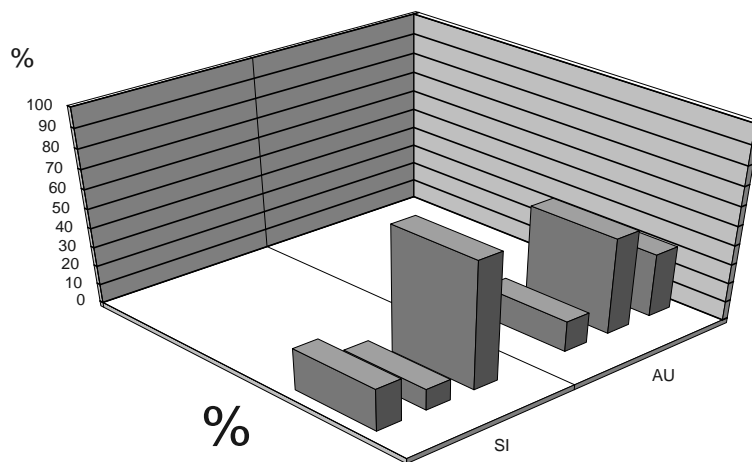


AE

Diagramm 2a: Gipfelmuster bei simuliertem (SI) und authentischem (AU) Material.

### GIPFELMUSTER (VS)

*mittel - spät - früh*

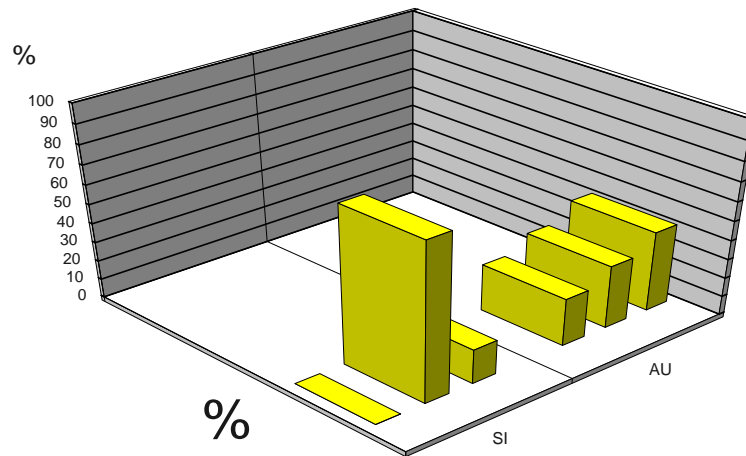


VS

Diagramm 2b: Gipfelmuster bei simuliertem (SI) und authentischem (AU) Material.

## GIPFELMUSTER (VA)

*mittel - spät - früh*



VA

*Diagramm 2c: Gipfelmuster bei simuliertem (SI) und authentischem (AU) Material.*

Diese Unterschiede können derzeit noch nicht interpretiert werden, weil für eine Interpretation die pragmatische Funktion von Gipfelmustern von Bedeutung ist. Eine Analyse einer Diskussion hat ergeben, daß die Realisierung früher Gipfelmuster Unterbrechungen verhindert, während die Realisierung später Gipfelmuster Unterbrechungen fördert (Moosmüller, in diesem Band). Interpretiert man nun diese Ergebnisse vorsichtig, so drücken frühe Gipfelmuster Bestimmtheit aus (analog zu Schmidts (1994) „Hörerreaktion nicht erwartet“), während späte Gipfelmuster dem Gesagten einen unsicheren Ton verleihen (analog zu Schmidts „Hörerreaktion erwartet“).

Folgende Fragen bleiben aber offen, bevor eine Interpretation der beobachteten Gipfelmuster erfolgen kann:

1. Gibt es einen Unterschied zwischen Hochsprache und Dialekt in der Realisierung von Gipfelmustern?
2. Wie werden Gipfelmuster funktional-pragmatisch in beiden Systemen eingesetzt.

### 3.3. Kontrastmaximierung

Unterschiede in der artikulatorischen Genauigkeit bei verschiedenen emotionalen Zuständen werden in vielen Untersuchungen beschrieben, so auch bei Trojan (1975), (vgl. Bezooyen 1984 für einen Literaturüberblick). Bezooyen (1984) weist diesem Parameter perzeptorisch allerdings geringe Bedeutung zu. Dennoch erscheint es uns wichtig, den Unterschied in der artikulatorischen Genauigkeit zwischen den unterschiedlichen emotionalen Zuständen zu untersuchen, und zwar unter dem Aspekt der Kontrastmaximierung.

Segmentale Prozesse (Entdeutigungs- und Verdeutlichungsprozesse, vgl. Dressler 1984) hängen von der postlexikalischen Betonung ab; wie Madelska & Dressler (1994) anhand des Polnischen und Tschechischen und Vollmann (1994) anhand der Wiener Monophthongierung gezeigt haben, kommt es bei Reduktionsprozessen zu einem „flattening of the metrical tree“, d.h., sekundärbetonte Silben werden zu schwachen Silben, schwache Silben werden stark reduziert, und schwächste Silben werden getilgt; Entdeutigungsprozesse werden so anhand universeller Präferenzen (der Binarität, des figure-ground-Kontrasts) vorhersagbar, da eine Schwächung von z.B. ssws zu swww im Sinne der Dresslerschen Natürlichkeitstheorie (vgl. Madelska & Dressler 1990) zu „unnatürlichen“ metrischen Strukturen führt, die durch

Reduktionen und Tilgungen wieder „natürlicher“ im Sinne der universellen Präferenzen werden (z.B. zu sww(w)). In einem Baummodell der metrischen Struktur einer Äußerungseinheit<sup>5</sup>, bedeutet dies ein „Verflachen“ der Struktur.

Die „postlexikalische“ Betonung wird, basierend auf der lexikalischen, von psychosoziofonologischen Parametern bestimmt (u.a. von den psychosozialen Gestimmtheiten der AE, VA und VS).

Während in casual speech weitgehend Entdeutlichungsprozesse untersucht wurden, handelt es sich beim vorliegenden Material um eine Untersuchung fortitiver Mechanismen; dabei ergab sich, daß die Emphase (psychosoziale „Verdeutlichung“) bei VS und VA ausschließlich durch phonologische Stärkungsprozesse ausgedrückt wird, bei AE hingegen zusätzlich auch durch phonologische Schwächungs-, Reduktions- und Tilgungsprozesse in unbetonten Positionen. Da es sich in dem vorliegenden Material nicht um casual speech handelt und daher theoriegemäß Entdeutlichungsprozesse nicht zu erwarten sind, müssen sie hier (im Sinne der universellen Präferenz für figure-ground-Kontrast) als Mittel der Kontrastmaximierung, d.h., als Hervorhebung der jeweiligen betonten Position verstanden werden.

### 3.3.1. Vereinigungsstreben (VS)

#### simuliert:

[b5çs}d5u:A:BE}li:p (44)<sup>6</sup> }kOm}g5i:bmçE daEnEé}hA:ndö\$ñ}lAsdEaEnEn}kös}ge:bm1]

```

postlex. : s   s w w   s   s   s   w   w w   s   w
lexikal. : s   s s w   s   s   s   w   s w   s   w
phon. Tr.: b5çs   }d5u:A:BE   }li:p   }kOm   }g5i:b   mçE   daEnEé}hA:nd
ö$ñ

```

Text : **B**ist du aber lieb! Komm, gib mir deine Hand und

```

postlex.: s   w   w w   s   w w
lexikal.: s   w   s w   s   s w
phon. Tr. }lAs   dE   aEnEn }kös }ge:bm
Text:   laß dir einen Kuß geben!

```

Bei diesem Affekt sind kaum Schwächungs-, Reduktions- oder Tilgungsprozesse zu beobachten. In der ersten Äußerungseinheit („bist du aber lieb“) kommt es zu einer Umstrukturierung von sssws zu sswws, wodurch die zweite Silbe von „aber“ stärker reduziert werden muß (Spirantisierung von /b/ und Reduktion von üblichem [ʔ] zu [E]). Die Reduktion des [ʔ] zu [E] betrifft auch die Dauer; sie beträgt 73 ms, gegenüber z.B. 166 ms in betontem [u:] von „du“. Die Reduktion von [ʔ] zu [E] läßt sich auch in „mir“ und „dir“ beobachten; bei „dir“ ist darüberhinaus ein Dialektalismus feststellbar: /dir/ → [d5ʔ] → [dE], der durch den folgenden anlautenden Diphthong begünstigt wird (Assimilation des [çÇ] an den Diphthong). „Dir einen“ ist stärker reduziert als „mir deine“; die Veränderung des F<sub>2</sub> gegenüber der Anfangsposition innerhalb des Diphthongs /ae&/ in „deine“ beträgt 78 Hz bei einer Dauer von 113 ms, in „einen“ lediglich 8 Hz bei einer Dauer von 54 ms. Beide Positionen können lexikalisch als sekundärbetont gelten; postlexikalisch aber ist „einen“ wesentlich stärker reduziert, was v.a. darauf zurückzuführen ist, daß es sich im Gegensatz zum Possessivpronomen um ein (typischerweise reduzierbares) Funktionswort handelt (funktional determinierte Prozeßapplikation, vgl. Fenk-Oczlon 1989). Darüberhinaus verfügt das Possessivpronomen über keinen auslautenden Nasal, wodurch mehr Zeit für die Artikulation der vorausgehenden Segmente vorhanden ist. Die Gesamtdauer von „einen“ und „deine“ ist hingegen fast gleich und beträgt bei „deine“ 284 ms, bei „einen“ 271 ms.

Hingegen sind in starken Positionen Verdeutlichungsprozesse zu beobachten, die sich auf die Dauer der Vokale auswirken. So beträgt die Dauer von [u:] in „du“ 166 ms, von [i:] in „lieb“ 258 ms, von [i:] in „gib“ 149 ms, von [a] in „Hand“ 136 ms, von [ö] in „Kuß“ 113 ms und von [e:] in „geben“ 287 ms. Die Dauerwerte liegen aber unter den gemessenen Werten beim simulierten Material von VA.

#### authentisch:

<sup>5</sup> Der Terminus „Satz“ muß hier vermieden werden.

<sup>6</sup> Die Zahlenwerte in Klammer beziehen sich auf die Dauer der Pausen (in ms).

[]vé: (302) d5Es}hOBE dE}e:S"ksOkt] (was denn?)  
 [d5"s}du:méEn}hE"tsb5lAt}b5çst] (na, des sogst mir zum ersten Mal)  
 [EnO}sOçd"s}çe:ts (1495) Z5d5esd5i}mandu}bçst}a:fOxd5i}fRü:O]<sup>7</sup>

postlex. : s w s w w s w s  
 lexikal. : s w s w s s w s  
 phon. Tr.: }vé: d5Es hOBE dE }e: S" ksOkt  
 Text : weil, das habe ich dir eh schon gesagt

postlex. : w s w s w s  
 lexikal. : w s w s w s  
 phon. Tr.: d5"s }du: méEn }hE"tsb5lAt }b5çst  
 Text : daß du mein Herzblatt bist

postlex. : 0 s w s w s s w s s  
 lexikal. : w s w s w s s w w s  
 phon. Tr.: Z5d5esd5 i}man du}bçst }a:fOx d5i }fRü:O  
 Text : verstehst, ich meine, du bist einfach die Frau

Auch beim authentischen Material kommt es kaum zu Schwächungs-, Reduktions- oder Tilgungsprozessen. In der ersten Äußerungseinheit („weil, das habe ich dir eh schon gesagt“) sind zwei Vokalassimilationsprozesse zu beobachten: [i:] (in „habe ich“) → [E], [d5" ] („dir“) → [dE], es handelt sich dabei um Assimilationen an das folgende [}e:] („eh“). Dialektales [So\$:] („schon“) wird zu [S"] reduziert. Aufgrund der kurzen Dauer des Vokals (35 ms) wird er zentralisiert und verliert auch die Nasalierung, d.h. der Nasal ist nicht mehr transparent<sup>8</sup>. In der zweiten Äußerungseinheit („daß du mein Herzblatt bist“) sind weder Schwächungs-, Reduktions- oder Tilgungsprozesse feststellbar, es kommt auch zu keinen Umstrukturierungen des metrischen Baumes. „Mein“ wird sogar für eine dialektale Äußerung stark diphthongiert; die Veränderung des F<sub>2</sub> gegenüber der Anfangsposition innerhalb des Diphthongs /ae&/ beträgt 75 Hz bei einer Dauer von 88 ms<sup>9</sup>. Auffällig ist in der dritten Äußerungseinheit („verstehst, ich meine, du bist einfach die Frau“) die Umstrukturierung von swws zu swss in „einfach die Frau“, d.h. der bestimmte Artikel wird durch Längung des Vokals (87 ms) zusätzlich gestärkt. Wie auch beim simulierten Material kommt es teilweise zu Stärkungsprozessen durch Längung der Vokale. So beträgt die Dauer von [é:] in „weil“ 308 ms, von [e:] in „eh“ 141 ms, von [O] in „gesagt“ 118 ms, von [u:] in „du“ 244 ms, von [Eç] in „Herzblatt“ 156 ms, von [e:] in „verstehst“ 110 ms, von [a] in „ich meine“ 42 ms (!), von [a:] in „einfach“ 106 ms und von [ü:O] in „Frau“ 269 ms.

Es kommt also bei beiden Sprechern zu keiner Kontrastmaximierung durch Entdeutlichung der unbetonten und Verdeutlichung der betonten Positionen, die Struktur des metrischen Baumes ist relativ ausgeglichen und weist im wesentlichen die Struktur swswsw... auf.

### 3.3.2. Verstandesmäßiger Ausdruck (VA)

simuliert:

[aEnE}AusA:gEdesd5i}grë:sn1}a:ind1}b1e:aEnAnE}glaECsçn}nEnEnvçaEnEi}glAiCi\$:nd (92)  
 }a:ind5}b5e:iErE}said5n1od"}thailE]

postlex. : w w s w w w w s w s w s w w w  
 lexikal. : s w s w w s w s w s w s w s w

<sup>7</sup> Die Äußerungseinheit „na, dann sag ich dir's jetzt“ wurde hier nicht analysiert, da sie eher als ungeduldige Interjektion zu verstehen ist (die angesprochene Person geht nicht auf das Werbungsverhalten ein), denn als zum eigentlichen Werbungsverhalten gehörig. Aus diesem Grund sind hier auch sehr viele Entdeutlichungsprozesse zu beobachten.

<sup>8</sup> Die unreduzierte dialektale Form der ersten Äußerungseinheit würde lauten: []vé: d5Es-}hOBi:dç}e:So\$:ksOkt]

<sup>9</sup> Im Vergleich dazu: die Veränderung des F<sub>2</sub> gegenüber der Anfangsposition innerhalb des Diphthongs /ae&/ in [}vé:] („weil“) beträgt 55 Hz bei einer Dauer von 308 ms! Die unreduzierte dialektale Form ist ein Monophthong mit Ersatzdehnung (siehe Moosmüller 1991, Vollmann 1991).



phon. Tr.: aEnE }AusA:gE dés d5i }grè:sn1 }a: ìnd1 }b1e: aEnAnE  
 Text: Eine Aussage, daß die Größen a und b einander

postlex. : s w s w w w w s 0 w  
 lexikal. : s w s w s s w s w w

phon. Tr.: }glaEC sçn }nEnEn vç aEnEi }glaIC i\$:nd  
 Text: gleich sind, nennen wir eine Gleichung und

postlex. : s w s w w s w w w s w  
 lexikal. : s w s s w s w s w s w

phon. Tr.: }a: ìnd5}b5e: iErE }said5n1 od" }thailE  
 Text: a und b ihre Seiten oder Teile.

Es kommt nur zu einer geringfügigen Umstrukturierung des metrischen Baumes und folglich auch zu wenigen Entdeutlichungsprozessen. In der ersten Äußerungseinheit („eine Aussage, daß die Größen a und b einander gleich sind“) kommt es lediglich zu zwei Tilgungen von Plosiven (in „einander“ [aEnAnE] und „sind“ [sçn]) sowie aufgrund der Schwächung zur Hebung des Vokals [a] zu [é] in „daß“ (regressive Vokalassimilation an das vorhergehende [E] in „Aussage“). In „einander“ wird der ursprüngliche auslautende [Ç] zu [E] reduziert. In Äußerungseinheit 2 („nennen wir eine Gleichung und“) können zwei Absorptionen beobachtet werden: [Ç] von „wir“ (unreduziert [vçÇ]) wird vom folgenden [aE] von „eine“ absorbiert (begünstigt durch die Aueinanderfolge von vier schwachen Silben!) und [iN] von „Gleichung“ wird mit vorhergehender Tilgung des Nasals vom darauffolgenden [i] von „und“ absorbiert. Dabei kommt es zu Ersatzdehnung des [i] (187 ms), die Nasalierung bleibt erhalten<sup>10</sup>. In Äußerungseinheit 3 („a und b ihre Seiten oder Teile) sind keine Entdeutlichungsprozesse zu beobachten.

Auffällig sind hingegen die vielen Verdeutlichungsprozesse, die in erster Linie die Dauer der Vokale und Diphthonge in starker Position sowie diphthongische Qualität letzterer betreffen<sup>11</sup> (siehe Tabelle 3):

Wort	DIF <sub>2</sub> (Hz) <sup>12</sup>	Dauer (ms)	Position
eine	320	80	schwach
Aussage	-305	240	stark
einander	47	114	schwach
gleich	648	204	stark
eine	54	66	schwach
Gleichung	688	275	stark
Seiten	946	257	stark
Teile	984	278	stark

Tab. 3: Differenz des F<sub>2</sub> der Endposition gegenüber der Anfangsposition und Dauer der Diphthonge.

Vokale in starker Position weisen ähnlich hohe Dauerwerte auf: [ë:] in „Größen“ 180 ms, [a:] in a<sub>1</sub> 255 ms, [e:] in b<sub>1</sub> 147 ms, [a:] in a<sub>2</sub> 212 ms und [e:] in b<sub>2</sub> 171 ms. Auch der Plosiv in „Teile“ wird gestärkt, die Dauer von Verschluss-, Plosions- und Aspirationsphase beträgt 202 ms.

Der Sprecher weist also eine Reihe von Verdeutlichungsprozessen auf, ohne gleichzeitig in schwachen Positionen Entdeutlichungsprozesse durchzuführen.

**authentisch:**

[n1dEs}vA:sjE}i:nEtvOs}d5emEd"nO\$N}kho:b5d5O\$m}fo:"hE"]  
 [na: (226) }hE:}h"i"x"mEi:}m"sd5En}hABORA}i:N1tv"i'd"}g"Ei\$N"}mo:AN1]

<sup>10</sup> Diese für den Sprecher unübliche Reduktion ist möglicherweise auf notwendiges Einatmen zurückzuführen, man beachte die darauffolgende kurze Pause von 92 ms.

<sup>11</sup> In Vollmann (1994) wurde gezeigt, daß in der Spontansprache der Wiener Hochsprache die F<sub>2</sub>-Veränderung in betonter Position durchschnittlich (nur) 100 bis 120 Hz ausmacht; lediglich bei isoliert gesprochenen Wörtern werden vergleichbare Werte gemessen (vgl. livonen (1989)).

<sup>12</sup> Differenz des F<sub>2</sub> der Endposition gegenüber der Anfangsposition.

```

postlex. : w w s w s w w s w w w 0
lexikal. : w s s w s s w s s w s s w s w
phon. Tr.: n1 dEs }vA:s jE }i: nEt vOs }d5e mEd~nO$N
Text: und das weiß ja ich nicht, was die miteinander

postlex. : s w s w s s s w w w
lexikal. : s s s w s s s w s s
phon. Tr.: }kho:b5d 5O$m }fo:"hE" na: }hE: }hĩ"x "mE i:
Text: gehabt haben vorher. Na, he, horch einmal ich

postlex. : s w s w w s w w w s w s w
lexikal. : s w s w w s w w w s w s w
phon. Tr.: }mĩ"s d5En }hABORA }i:N1tvĩ d"}g"Ei$N" }mo:AN
Text: muß den Hawerer irgendwo derglengen morgen.

```

Auch beim authentischen Material kommt es nur zu geringfügigen Umstrukturierungen des metrischen Baumes und folglich zu wenigen Entdeutlichungsprozessen. In der ersten Äußerungseinheit („und das weiß ja ich nicht“) wird „und“ zu einem silbischen Nasal gekürzt, ursprüngliches [O] von „ja“ wird progressiv an [i:] „ich“ assimiliert und folglich zu [E] gehoben. In Äußerungseinheit 2 („was die miteinander gehabt haben vorher“) kommt es zu stärkeren Umstrukturierungen: swsws („miteinander gehabt“<sup>13</sup>) wird zu www0s. Dadurch, daß in „miteinander“ beide starken Positionen geschwächt werden, muß die Endsilbe getilgt werden, mit folgender Assimilation des Nasal an den velaren Plosiv. [ç] aus „miteinander“ wird an den Folgevokal assimiliert und folglich zu [E] gesenkt. [h] aus „haben“ wird getilgt bzw. von der vorhergehenden geringfügigen Aspiration des Plosivs absorbiert. In Äußerungseinheit 3 („na, he, horch einmal“) kommt es zur Umstrukturierung und folglich zur Reduktion von „einmal“. Die ursprünglich starke Silbe [mOe] „mal“ wird zu [mE] reduziert, damit die Dauer von 166 ms des ebenfalls geschwächten „ich“ aufrechterhalten werden kann. In Äußerungseinheit 4 („ich muß den Hawerer irgendwo derglengen morgen“) wird intendiertes [A] in „Hawerer“, bedingt durch die Lippenrundung des bilabialen Frikativs, ebenfalls gerundet und somit zu [O]. [Ç] von „irgendwo“ wird getilgt, mit einhergehender Ersatzdehnung (Dauer des [i:] beträgt 136 ms). In „wo“ wird [o] and die Höhe des vorhergehenden Vokals [i:] assimiliert, es kommt zu einer Reduktion der Dauer (30 ms), wodurch der Vokal auch seine Spannung verliert.

Wieder wirken sich Verstärkungsprozesse auf die Dauer der insbesondere betonten Vokale aus. Die Dauer von [A:] aus „weiß“ beträgt 210 ms, von [ç] aus „ich<sub>1</sub>“ beträgt 135 ms, von [o:] aus „gehabt“ 132 ms, gleichzeitig wird der Vokal gespannt und der velare Plosiv aspiriert, von [e:] aus „he“ 130 ms, von [Ei\$] aus „derglengen“ 141 ms, zusätzlich wird der Vokal diphthongiert, dabei aber der Lateral bei gleichzeitiger Palatalisierung des velaren Plosivs getilgt (Kontrastmaximierung innerhalb einer Silbe), von [o:A] aus „morgen“ 119 ms, zusätzlich wird der Vokal gespannt artikuliert.

Gegenüber dem simulierten Material sind die Verstärkungsprozesse allerdings weniger ausgeprägt (die Dauern der Vokale sind kürzer)<sup>14</sup>, Stärkungs- und Entdeutlichungsprozesse sind relativ ausgeglichen, während es beim simulierten Material zu einer Überbetonung der Verstärkungsprozesse kommt, die der Äußerung einen artifiziellen Ausdruck verleihen.

### 3.3.3. Aggressive Ergotropie (AE)

#### simuliert:

```

[nOx}ain}vORd̄indi:C}hAOinEnEnE}OrféEg"E"ri$nd"si: (36) }i$Mf"SEi$dé}mEi$N$S̄i:}
[AézéEnE}frECéEdçsmEdOVEméEnEm}le:nOxni$:Cd5}O$ndEgekOmén]

```

```

postlex. : w s s w w s w w w w s w 0 w s w w
lexikal. : w w s w s s s w s w s s w w s w s

```

<sup>13</sup> die unreduzierte dialektale Form würde lauten: [mçdAnündÇkho:bd]. Vgl auch „Gleichung und“ des simulierten Materials.

<sup>14</sup> livonen (1987) hat festgestellt, daß die Vokaldauern des norddeutschen Sprachraums gegenüber dem süddeutschen Sprachraum kürzer sind. Innerhalb der österreichischen Varietät sind die Vokaldauern der hochsprachlichen Varietät gegenüber den dialektalen Varietäten kürzer (Moosmüller 1991)

phon. Tr.: nOx }ain }vORd ind i:C }hAO inEn EnE }OrféEg E""ri\$nd"  
si:

Text: Noch ein Wort und ich hau Ihnen eine Ohrfeige herunter, Sie

postlex. : s w w w s w 0 w w s w w w w  
lexikal. : s w s w s s w s w s w s w s  
phon. Tr.: }i\$Mf"SEi\$dé }mEi\$N\$ s<sup>i</sup>: Aéz éEnE }frECéEd çs mE  
dOV

Text: unverschämter Mensch, Sie! So eine Frechheit ist mir doch

postlex. : w w w s 0 w w s w w s w  
lexikal. : w s w s w s w s w w s w  
phon. Tr.: E méEnEm }le: nOx ni\$:Cd5 }O\$ndEgekOmén

Text: in meinem Leben noch nicht untergekommen!

Bei diesem Affekt kommt es neben der emphatischen Verdeutlichung der Schlüsselwörter durch postlexikalische Nicht-Verstärkungen der übrigen lexikalisch unbetonten Positionen zu einer Anhäufung vieler (maximal sieben) unbetonter Silben in einem Fuß und daher auch zu vielen Schwächungs-, Reduktions- und Tilgungsprozessen – sogar bei emphatischer Rede. In der ersten Äußerungseinheit („noch ein Wort und ich hau Ihnen eine Ohrfeige herunter“) überwiegen noch die Verdeutlichungsprozesse gegenüber den Entdeutlichungsprozessen, zu einer Umstrukturierung kommt es lediglich in „Ohrfeige herunter“, nach Tilgung des [h] wird die Endsilbe [E] von „Ohrfeige“ vom nun anlautenden [E] absorbiert. Dadurch wird die metrische Struktur swwwsww wieder zu zwei Daktylen (swswsw) reduziert. In „eine“ wird der Diphthong insbesondere qualitativ reduziert. Beträgt im unbestimmten Artikel in VS die Differenz von F<sub>2</sub> immerhin noch 8 Hz und in VA sogar noch 54 Hz, wird nun der Diphthong zur Gänze monophthongiert und zusätzlich an den Folgevokal assimiliert ([E] statt [é]). Dies läßt bereits auf eine Kontrastmaximierung schließen. In der zweiten Äußerungseinheit („Sie unverschämter Mensch Sie!“) kommt es ebenfalls zu Umstrukturierung: swsw wird zu swww in „unverschämter“. Dabei wird der Nasal getilgt, der vorhergehende Vokal wird dabei aber diphthongiert mit einhergehender Nasalisierung des zweiten diphthongischen Elements. Der intendierte [ç] der Endsilbe wird an den vorhergehenden Diphthong assimiliert und somit zu [é] gehoben. Es handelt sich in diesen beiden Äußerungen somit in erster Linie um qualitative Reduktionen. In Äußerungseinheit 3 („so eine Frechheit ist mir doch in meinem Leben noch nicht untergekommen“) kommt es zusätzlich zu quantitativen Reduktionen: In „so“ wird der Vokal getilgt und der Frikativ lenisiert, in „Frechheit“ kommt es zur Tilgung des [h], die Veränderung des F<sub>2</sub> im Diphthong beträgt lediglich 55 Hz (gegenüber noch 100 Hz in „eine“), in „ist“ kommt es zur Tilgung des Plosivs, intendiertes [mçç] „mir“ wird zu [mE] reduziert, diese Reduktion beinhaltet gleichzeitig eine Assimilation an den vorhergehenden Vokal, intendiertes [x] in „doch“ wird aufgrund der stimmhaften Umgebung ebenfalls stimmhaft, bei „in“ kommt es zur Tilgung bzw. Absorption des Nasals mit einhergehender Vokalassimilation, intendiertes [ç] wird aufgrund der Umgebung zu [E] gesenkt, der Diphthong in „meinem“ weist eine Veränderung des F<sub>2</sub> von lediglich 39 Hz auf, in „Leben“ kommt es zur Tilgung einer gesamten Silbe, wodurch die Dauer des betonten Vokals (170 ms) stärker hervorgehoben wird. Durch diese Tilgung kann die Struktur von swwwswww zu swswsww umgeformt werden.

Diesen Reduktionsprozessen stehen Verdeutlichungsprozesse gegenüber, die allerdings in den ersten beiden Äußerungseinheiten stärker sind. In der ersten Äußerungseinheit fällt vor allen Dingen die Unterdrückung der r-Vokalisierung in „Wort“ und „Ohrfeige“ auf, die Dauer der Vibranten beträgt 60 ms bzw. 43 ms<sup>15</sup>. Die Veränderung des F<sub>2</sub> beträgt in dem gestärkten unbestimmten Artikel „ein“ 859 Hz. In Äußerungseinheit 2 kommt es zur Diphthongierung des [E] in „Mensch“ mit einer Dauer von 170 ms. In Äußerungseinheit 3 kommt es in „Frechheit“ zu einer Längung der gesamten betonten Silbe; die Dauer des Frikativs beträgt 138 ms, die des Vibranten 89 ms und die des Vokals [E] 193 ms.

#### authentisch:

[e:}Ru:fsédanA}OEm1}Sd5OĩmbéEidi"}O\$:E:man (181) b5çs}d5E:p<sup>h</sup>O\o\O"}O:s]

<sup>15</sup> Die Unterdrückung der r-Vokalisierung ist für die Österreichische Hochsprache unüblich und somit als Verdeutlichung mit bühnensprachlichen Mitteln zu werten (vgl. auch die Verstärkung mit hochsprachlichen Mitteln im authentischen Material; dialektales [Riçf] wird zu hochsprachlichem [Ru:f]).

```

postlex. : w s w w w s w s w w s
lexikal. : s s s s w s w s s s s
phon. Tr.: e: }Ru:f séd anA }OEm1 }Sd5Oim béEi di" }O$:
Text: ich ruf seit einer halben Stunde bei dir an

postlex. : s 0 w w s w w w s
lexikal. : s w s s s w s w s
phon. Tr.: E: man b5çs }d5E:p^hO\ o^O" }O:s
Text: he ich meine, bist deppert oder was

```

Die Kontrastmaximierung von figure und ground ist beim authentischen Material eindrucksvoller; Reduktionsprozesse stehen meistens unmittelbar mit einer Hervorhebung der betonten Silbe oder des betonten Vokals in Zusammenhang. In „halben“ kommt es zur Tilgung des [h], wodurch die Betonung sofort auf den Vokal fällt. In „he ich meine“ kommt es zur Tilgung einer Silbe („ich“), wodurch die Struktur von swwws in swws umgeformt werden kann. In „deppert oder was“ kommt es zwar zu keiner Restrukturierung, doch werden die unbetonten Silben zugunsten des betonten „depp“ erheblich qualitativ und quantitativ reduziert. Es kommt zur Vokalassimilation des intendierten [A] aus „deppert“ an den Folgevokal [o], die beiden Plosive werden ebenfalls zu einem alveolaren flap reduziert. In der betonten Silbe kommt es dagegen zur Dehnung des [E] (161 ms), sowie zur für den Dialekt unüblichen Aspiration des Plosivs (Dauer von Verschuß-, Plosions- und Aspirationsphase 152 ms). In „was“ wird wieder der schwer perzipierbare Frikativ zugunsten des Vokals getilgt, der Vokal wird erheblich gedehnt (297 ms).

Weitere auffällige Verstärkungsprozesse sind die hochsprachliche Artikulation von „rufe“ ([Ru:f] statt dialektal [RiÇf]), die vorhergehende Silbe „ich“ wird qualitativ reduziert ([E:] statt [i:]), die Diphthongierung von [i] in „Stunde“, mit einhergehender Tilgung des auslautenden Plosivs und der Assimilation des Nasals an den folgenden Plosiv, die Dehnung des Vokals „an“ (234 ms).

Zusammenfassend läßt sich sagen, daß bei VS und VA kaum Kontrastmaximierung von figure und ground zu beobachten ist, während bei AE diese am stärksten hervortritt. Beim simulierten Material ist eine geringere Ausgeglichenheit zwischen betonten und unbetonten Positionen feststellbar, die sich in einer Maximierung von Verdeutlichungsprozessen gegenüber geringerem Auftreten von Entdeutlichungsprozessen zeigt (besonders bei VA). Dies ist beim authentischen Material nicht in diesem Ausmaß der Fall. Bei AE ist die Kontrastmaximierung beim authentischen Material stärker, da Verdeutlichungsprozesse meist zuungunsten der unbetonten Silben wirken, während beim simulierten Material (besonders bei Äußerungseinheit 1 und 2) Verstärkungsprozesse ohne erhebliche Reduktionsprozesse auftreten.

#### 4. Literatur

**Bezooyen, Renee van (1984):** Characteristics and Recognizability of Vocal Expressions of Emotions. Dordrecht: Foris Publications.

**Deutsch, Werner A. & Anton Noll (1990):** S\_Tools. The Integrated Workstation for Music, Acoustics, Speech, and Signal Processing. Programmers Reference Manual. Vienna: Austrian Academy of Sciences, Acoustic Research Department.

**Deutsch, Werner A. & Anton Noll (1994):** S\_Tools. Datenerfassung, Speicherung und Digitale Signalverarbeitung für Akustik, Lärm, Sprache und Musik. Vienna: Austrian Academy of Sciences, Acoustic Research Laboratory.

**Dressler, Wolfgang U. (1984):** xxxx. in: Phonology Yearbook 1

**Fenk-Oczlon, Gertraud (1989):** Geläufigkeit als Determinante von phonologischen Backgrounding-Prozessen. In: Papiere zur Linguistik 38: 91-103.

**Gartenberg, Robert & Christa Panzlaff-Reuter (1991):** Production and Perception of F<sub>0</sub> Peak Patterns in German. in: Arbeitsberichte des Instituts für

- Phonetik und digitale Sprachverarbeitung der Universität Kiel 25: 29-115. livonen, Antti (1987): Monophthonge des Gehobenen Wienerdeutsch. In: Folia Linguistica 21, 293-336.
- livonen, Antti (1987):** Monophthonge des Gehobenen Wienerdeutsch. In: Folia Linguistica 21, 293-336.
- livonen, Antti (1989):** Regional German Vowel Studies. Helsinki (= Helsingin Yliopiston Fonetikan Laitoksen Moniteista. Mimeographed Series of the Department of Phonetics, University of Helsinki. Nr.15).
- Kohler, Klaus J. & Robert Gartenberg (1991):** The Perception of Accents: F<sub>0</sub> Peak Height versus F<sub>0</sub> Peak Position. in: Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel 25: 219-243.
- Madelska, Liliana & Wolfgang U. Dressler (1994):** Postlexical Stress Processes and Their Segmental Consequences Illustrated with Polish and Czech. in: Bernhard Hurch (ed.): Natural Phonology. Workshop at the Annual Meeting of the Societas Linguisticae Europaeae in Bern. Wuppertal.
- Markel, John D. & Augustine H. jr. Gray, (1976):** Linear Prediction of Speech. Berlin: Springer.
- Moosmüller, Sylvia (1991):** Hochsprache und Dialekt in Österreich. Soziophonologische Untersuchungen zu ihrer Abgrenzung in Wien, Graz, Salzburg und Innsbruck. Wien: Böhlau (= Sprachwissenschaftliche Reihe. 1).
- Moosmüller, Sylvia (in diesem Band):** The Relevance of F<sub>0</sub> in Discussions and its Effects on Female/Male Interaction Patterns.
- Moosmüller, Sylvia & Ralf Vollmann (1994):** Dialekt- und Hochsprachevariation bei Kleinkindern in Wien: Phonologie. in: Harald Burger & Annelies Häcki-Buhofer (eds.): Spracherwerb im Spannungsfeld von Dialekt und Hochsprache. Bern: Lang (= Züricher Germanistische Studien 38), pp. 109-128.
- Schmidt, Jürgen E. (1994):** Grammatik und Pragmatik der Intonation im Standarddeutschen. Vorstellung eines Forschungsprojekts. Vortrag gehalten anlässlich der 25. Jahrestagung der Gesellschaft für Angewandte Linguistik (Sektion Phonetik) an der Universität Trier, 21.-23.9.1994.
- Trojan, Felix (1975):** Biophonetik. Zürich: B.I. – Wissenschaftsverlag (hrsg. von Herbert Schendl).
- Vollmann, Ralf (1991):** Die Wiener Monophthongierung. Diplomarbeit, Wien.
- Vollmann, Ralf (1994):** Phonetics of informal speech: The Viennese Monophthongization. In: Studia Phonetica Posnaniensia 5: 1-15.
- Williams, Carl E. & Kenneth N. Stevens (1972):** Emotions and Speech: Some Acoustical Correlates. In: Journal of the Acoustical Society of America 52: 1238-1250.