

# **VOWELS IN STANDARD AUSTRIAN GERMAN**

**An Acoustic-Phonetic and Phonological Analysis**

**VON  
DR. PHIL. SYLVIA MOOSMÜLLER**

Habilitationsschrift  
eingereicht an der  
Philologisch-Kulturwissenschaftlichen Fakultät  
der Universität Wien

Wien, im Mai 2007

Sylvia Moosmüller

*to my children Caroline and Bernd*



## ACKNOWLEDGMENTS

I would like to express my gratitude to the people and institutions that made this work possible:

The Acoustics Research Institute of the Austrian Academy of Sciences and the Institute of Linguistics at the University of Vienna, who provided me with the necessary resources to conduct my analyses.

Wolfgang U. Dressler accompanied, guided, and influenced my academic career for many years. With his detailed and meticulous comments he sharply pointed out argumentative weaknesses and inconsistencies and did not spare the effort to re-read any revision of my work.

Werner A. Deutsch was open for discussions at any time I needed to clarify critical points. He always watched attentively that I looked at theories and concepts from different angles.

The influence of Ruth Wodak, though less obvious perhaps, should, however, by no means be underestimated, since the sociolinguistic aspect of language is a cornerstone of this work.

All my colleagues at the Acoustics Research Institute, especially Timo Becker who, with his marvellous accuracy, had a highly critical eye on both content and form, and Jonathan White, who, known for his patience, transported my English into a readable and understandable form. Special thanks go also to the STx Software Group, Anton Noll, Jonathan White, and Christian Gottschall who were always ready to help me with whatever problem I had to deal with.

My husband Ralf Vollmann took over the thankless task of formatting this text. Moreover, he was ready to discuss all aspects not directly related to phonetics or phonology.

Special thanks go to the anonymous speakers who made, by their willingness to participate, this work possible.

My thanks go also to those who supported me with helpful tips, encouraging words, and useful discussions: Gelong Lodreu, Aurelia Brait-Roher, Angelika Braun, Ursula Doleschal, Karin Freisinger, Theodor Granser, Jens-Peter Köster, John Rennison, Magda Rzepecka, and, last but not least, my children Caroline and Bernd, to whom this work is dedicated.

## Table of Contents

VOWELS IN STANDARD AUSTRIAN GERMAN .....	1
An Acoustic-Phonetic and Phonological Analysis .....	1
ACKNOWLEDGMENTS .....	4
Table of Contents .....	6
1. Introduction: Setting the Theoretical Framework.....	1
1.1. The socio-pragmatic foundation of phonology and phonetics.....	3
1.2. The phonological system .....	9
1.3 Standard Austrian German .....	15
2. Methods and Data.....	18
2.1. Speakers and Material .....	18
2.2. Measurements.....	20
3. Interpretation of formant measurements.....	29
3.1. The traditional F1/F2 representation .....	29
3.2. Constriction locations for vowels: discrete or gradual?.....	39
3.3. The acoustic-articulatory relationship .....	43
4. Vowel inventory and features.....	51
4.1. The feature [±tense]: brief historical outline .....	54
4.2. Temporal Analysis.....	56
4.2.1. Isochrony in Standard German .....	56
4.2.2. Isochrony in Central Bavarian .....	59
4.2.3. Isochrony in Standard Austrian German .....	62
4.2.4. Vowel duration in Standard Austrian German.....	66
4.2.5. Duration and Laxness .....	72
4.3. Articulatory investigations on vowels .....	75
4.3.1 Articulatory settings for the opposition traditionally termed [± tense].....	75
4.3.2 The features [± front] and [± high].....	82
4.4. Acoustic analysis of the vowels of Standard Austrian German.....	85
4.4.1. The pre-palatal vowels.....	86
4.4.2. The mid-palatal vowels .....	96
4.4.3. The velar vowels.....	106
4.4.4. The upper pharyngeal vowels.....	109
4.4.5. The lower pharyngeal vowels.....	112
4.4.6. The role of F0 .....	115
4.4.7. Evaluation of the results .....	120
5. Coarticulation .....	122
5.1. What is coarticulation?.....	122
5.2. Articulatory vs. acoustic observations.....	136
5.3. Processes vs. Coarticulation .....	147
5.3.1. The vowel /i/.....	150
5.3.2. The vowel /a/.....	159
5.3.3. The vowels /u, o/ .....	166
5.4. Conclusion.....	170
6. Vowel and Vowel Variability.....	172
6.1. Target, invariance, and target undershoot.....	174
6.2. Undershoot vs. Processes .....	179
6.3. Phoneme, allophone, target, and processes .....	185
6.4. Missed targets.....	187
6.2.1 Speech errors .....	187
6.4.2 Articulatory avoidance .....	188

6.5. Another look at invariance .....	194
6.6. Prosodic analysis of variability.....	198
6.6.1. Duration.....	198
6.6.2. Stress .....	203
6.6.3. Secondary stress .....	208
6.6.3.1 The vowel /a/.....	210
6.6.3.2. The vowel /i/.....	212
6.6.3.3. The vowel /ɪ/.....	213
6.6.3.4. The vowel /ɛ/.....	215
6.6.3.5. The vowel /e/.....	216
6.6.3.6. The vowel /u/.....	217
6.6.3.7. Summary of the results on secondary stress .....	218
6.6.4. Sentence Stress.....	218
6.6.5 The rhythm of speech .....	223
7. Speaker-specific traits .....	234
8. Summary and Outlook.....	244
9. References .....	248
Appendix .....	269

Sylvia Moosmüller

# 1. Introduction: Setting the Theoretical Framework

Language serves two main functions: a cognitive and a communicative function. The functions of linguistic components are subordinate to these main functions (Dressler 2002: 92f). The work at issue deals with acoustic phonetics and phonology, whose functions are subordinate to the communicative function of language<sup>1</sup>.

The contribution of phonology and phonetics to serving the communicative function is not as straightforward as the contribution of the other linguistic components, e.g. discourse. In fact, these other components serve the communicative function in the most indirect way (Dressler 2002). This does not, of course, diminish the importance of phonetics and phonology for the communicative function. Within the framework of Natural Phonology, it is the role of prelexical phonological processes to “merge conceivable sounds into the phoneme inventory of each language” and “to govern the phonotactics of phonemes” (Dressler 1984: 30) of a given language or language variety<sup>2</sup>. In line with Baudouin the Courtenay, the phoneme is defined as sound intention. This implies that phonemes are “fully specified, because they could not be intended otherwise” (Dressler 1984: 32). The semiotic model elaborated by Dressler (1980, 1984, 1985) gave the basis for the distinctive function of phonemes and features (“the signantia of a sign should be distinguishable” 1984: 35). These language-specific fully specified and distinctive sound intentions and the language-specific phonotactics are, of course, pronounceable and perceivable for the speakers of a given language or

---

<sup>1</sup> Whether phonology serves the cognitive function as well is discussed in Dressler (2002).

<sup>2</sup> The phoneme inventories and, consequently, their sequential orders might differ within the varieties of one language, e.g. the phoneme inventory of Standard Austrian German differs from the phoneme inventory of Austrian dialects (see Dressler & Wodak 1982, Moosmüller 1987, 1991).

language variety. In principle, as soon as the phoneme inventory is established<sup>3</sup>, the speaker is able to execute the communicative function in – from a phonological point of view – the most reliable way, namely in form of a biunique relationship between phoneme and phonetic output, i.e. “between signans and signatum” (Dressler 1984: 36). However, in order to ensure correct perception of the intended phoneme sequence, various adaptations are necessary. For example, lip protrusion of a rounded vowel has to start in the preceding consonant in order to make protrusion audible (Maeda 1999, Vaxelaire et al. 2003). Since speech is planned in terms of auditory targets (Guenther et al. 2006, Perkell et al. 2006), such processes are absolutely essential and, consequently, obligatory. The majority of processes are, however, language-, language variety-, or even speaker-specific. Anticipatory lip protrusion, for example, might affect even more preceding segments than just the immediately preceding consonant (see 5.1, Moosmüller 2007b). However, by applying anticipatory lip protrusion, motor complexity is increased (Van der Merwe 1997, Theron 2003), because lip protrusion is not related to the preceding phonemes (provided these are specified as unrounded). Therefore, one might ask why speakers abandon the biunique relationship and apply – from a cross-linguistic perspective – all sorts of processes? Why do French speakers apply anticipatory lip protrusion and, by doing so, not only increase motor complexity, but also risk distinctiveness (e.g. *détour* vs. *de tour*)? Why do Standard Austrian German speakers prefer to restrict anticipatory lip protrusion to the immediately preceding consonant? Given the language-specific and often speaker-specific application of processes, the answer can hardly be found in either enhancement of pronounceability or enhancement of perceptibility. However, languages have preferences about which processes they apply under which conditions, and which they suppress. Phonology and phonetics, together with the remaining linguistic components, therefore play a decisive role in distinguishing languages or language varieties from

---

<sup>3</sup> I.e., first language acquisition is completed and the speaker has no speaking or hearing disorders.

each other. The smaller the sociological entity of speakers, the greater the distinctive role of phonology and phonetics. A non-native speaker is immediately detected, especially by his or her inability to apply the proper language-specific processes. The same holds for speakers of different varieties. This distinguishing function is directly related to the communicative function.

The important role of phonetics and phonology in serving the communicative function of language becomes most obvious when the situational aspect is more explicitly emphasized. Studies both on sociophonology and sociophonetics unanimously prove that neither the frequency of occurrence of processes, nor the nature of the processes applied, is left to chance. The phonology and phonetics of a language, or of a variety, informs the listener to a large extent, not only about the regional and the social background of a given speaker, his or her age and sex, but also about the emotional state, the attitude of the speaker towards the listener(s), and his or her discursive intentions. Therefore, both the application and the non-application of a process have implications, and consequently, phonetics and phonology have to be conceptualised within a wider framework.

## **1.1. The socio-pragmatic foundation of phonology and phonetics**

Language is used by human beings to organize, maintain, or change their social life (both via communication and cognition). Therefore, “it is best described and understood as a system of goal-directed actions within its social frame (Dressler & Moosmüller 1991: 136). Most, if not all, investigations on speech behaviour came to the conclusion that speech behaviour differs according to the situation in which a specific interaction takes place. Speakers talk differently in formal speech situations and casual speech situations. It is assumed that in casual speech situations, speakers can feel more at ease and therefore are allowed to exert less effort while speaking; the result of this is less

clear speech. Implicitly, such a top-down analysis assumes that with an increasing informality of the situation, speakers gradually approach the rest position. This would further imply that in an informal speech situation, articulators are moved with less effort and removed from the neutral position as little as possible. Therefore, with an increasing informality of the speech situation, more and more processes which ease articulation (assimilations, reduction, deletions) are applied, reflecting the speaker's wish to economize the movements of his or her articulators as often and as much as possible.

However, what would happen if a speaker "strained" him- or herself in a cosy get-together with friends and spoke as if he or she had to take an exam or to apply for a job? This person would be laughed at, i.e. the speech behaviour of this person would be sanctioned. Therefore, a casual speech situation does not evoke processes which ease articulation (less economy or cost), but it evokes "casual speech processes", i.e., processes that are expected and defined as adequate in a casual speech situation. Speakers know very well how to behave in diverse speech situations, and, in case they don't, e.g. if they are confronted with a situation they are not acquainted with, they feel insecure and mix up processes (e.g. politicians demonstrating the common touch.). Therefore, in a casual speech situation, a person speaks the way it is expected from her or the way she thinks it is expected from her, but not according to a principle of least effort. Speech behaviour, in any situation, is highly regulated by mostly unwritten norms.

The top-down approach is reflected in phonetic and phonological theories which assume that speech behaviour is an activity which has to balance the needs of the speaker and the needs of the listener. In this concept of balancing, however, the needs of the speaker are implicitly or explicitly<sup>4</sup> defined as trying to exert as little effort as

---

<sup>4</sup> Lindblom (1983, 1990) compares the activity of speaking with the activity of cleaning a window. However, the difference is evident for any sociolinguist: speaking is a direct social activity, whereas window cleaning is not (there is no social interaction between the cleaner and the window). Therefore, the act of speaking is conducted by other principles than the act of cleaning a window.

possible while speaking. Since, on the other hand, the hearer should understand what the speaker wants to say, the speaker has to find a middle way; the speaker has to balance his or her laziness with the willingness of the listener to understand what he or she is saying. In such a dichotomous concept, speaker and listener are conceived as antagonistic (“tug of war” – Lindblom 1990). The intention of the speaker, in this concept, is to deviate as little as possible from a neutral vocal tract configuration.

“[...] that unconstrained a movement tends to default to a low-cost form of behavior. Accordingly, when an /i/ is produced without a bite-block, a tongue gesture is invoked that deviates little from a neutral configuration (,economy’).“ (Lindblom 1990: 417)

Unfortunately, the speaker has to move the articulators, otherwise the other person would not understand what he or she wants to tell him or her. I will show in 6.4 that the result of such an attitude from the side of the speaker is by no means an increased application of processes of ease of articulation, but, on the contrary, a very unsystematic application of all sorts of processes. Moreover, articulatory analyses emphasize the high precision with which we move our articulators (Wood 1982, Mooshammer 1998, Hoole & Mooshammer 2002), a tug of war attitude on the side of the speaker could never result in high articulatory precision.

Therefore, a dichotomous concept, in which the speaker’s needs are in contradiction to the needs of the listener, is wrong. The results of discourse analysis work vividly show that in speech situations, speakers are highly interested in being listened to, that they even interrupt others in order to speak themselves, that they try to draw attention to themselves, and that they want to be evaluated positively by the listeners, both as concerns content and form<sup>5</sup>.

Processes usually classified as ease of articulation are, in fact, responses to sequences excluded by prelexical processes (see 1.2), e.g. the widespread process of nasal assimilation, where the nasal consonant assimilates to the place of articulation of

---

<sup>5</sup> This is, of course, a highly shortened account of social interaction; the reader is referred to sociolinguistic literature for a thorough picture of how social interaction works in its multiple facets. The presentation suffices, however, to show that the speaker’s intention is not guided by a principle of least effort.

the following consonant (in English, for example, “im+possible”, Italian “im+possibile”, or German “un+möglich” (impossible) etc.). These sequences are caused by morphology. The “difficulty”, which has arisen through a prelexically excluded combination of phonemes (alveolar + bilabial), is eliminated by a prelexical process in English and Italian<sup>6</sup>, and by a postlexical process in German. A similar treatment can be observed for /s#f/ sequences in German, e.g. /hɑʊs#fʊɛ/ “Hausschuhe” (slippers). These processes aim at restoring a “pronounceable” sequence in the given language, whereby “pronounceability” is defined by the prelexical processes of a given language. These processes, which repair the sequential order of a language, are not restricted to informal speech situations.

The situation where the nasal and the preceding consonant are separated by a vowel, as in /tragen/ “tragen” (to carry) etc., is slightly different. In order to meet the same condition as above, the vowel has to be deleted first. In Standard Austrian German, vowel deletion is restricted to less formal speech situations. Consequently, nasal assimilation experiences the same situational restrictions<sup>7</sup>. But as soon as the vowel is deleted, the nasal has to be assimilated.

Casual speech processes therefore have to be kept apart from processes which repair disallowed phoneme combinations. The latter are independent of the speech situation and cannot be classified as casual speech processes. Casual speech processes serve other principles. In Austrian German, a typical process often seen as ease of articulation is the spirantization of lenis plosives in the intervocalic position. In the intervocalic position, /b, d, g/ can become /β, ð, γ/, as in [ɑ:βa] “aber” (but), [laɛða] “leider” (unfortunately), or [he:γe] “hege: 1<sup>st</sup> P.” (to take care). Since the air can flow

---

<sup>6</sup> Italian also has postlexical processes for sequences, e.g. “un bacio” (a kiss), which might even further assimilate to “um masu” in Sicilian (Hyman 2001). The same process might occur in English “non#basic”.

<sup>7</sup> In the Bavarian variants of German, including Standard Austrian German, vowel deletion between two nasal consonants (e.g. /ʃvɪmən/ “schwimmen” (to swim)) is suppressed. This corroborates the assumption that process application is not primarily lead by ease of articulation.

continuously in the case of the fricative, instead of being stopped during the occlusion phase, this process can be analyzed as assimilation to the surrounding vowels which also allow the air to flow continuously. However, there is no evidence that a fricative in the intervocalic position should be easier to produce than a plosive, or that assimilation itself eases articulation. Therefore, what happens by means of this process is a diminution of the figure-ground contrast (Dressler 1984, 1996), by signalling the interactional situation (casual) or the prosodic position (weak)<sup>8</sup>. The application of processes obeys semiotic (Dressler 1984, 1985, 1996) and socio-pragmatic (Dressler & Moosmüller 1991) principles other than an alleged low cost principle from the side of the speaker. Speech behaviour and variability, i.e. process application, are defined by the interactional situation, which again is defined by the social norms set up by the members of the speech community (see 1.3). The extent of the figure-ground contrast lies in the definition of a given interactional situation and only marginally in the sphere of influence of the speaker, who is allowed some individual latitude in process application. A speaker will at first assess a given speech situation and then adequately apply or suppress – as good as he or she can<sup>9</sup> – processes.

Output variability lies within the domain of postlexical process application or suppression. On the postlexical level, the phonology of a language provides its users with a set of phonological processes which have to be applied or suppressed, according to the demands of a given situation. In the same way as prelexical processes, which form the phoneme inventory of a given language, postlexical processes are universal in character, in the sense that they are “a latent invariable of human language” (Dressler 1979: 259), but language specific in application. Postlexical processes are phonetically motivated, in the sense that they either enlarge or diminish the figure-ground contrast in

---

<sup>8</sup> Adequately, Dressler (1984) substituted the terms „fortition“ and „lenition“ by foregrounding processes and backgrounding processes, respectively.

<sup>9</sup> This restriction refers to the speaker’s ability to deal with different speech situations. A speaker’s unwillingness to behave according to the norms of a given situation has to be considered and included as well (e.g. covert prestige).

a given interactional situation. The phonetic plausibility of postlexical processes has often been misinterpreted as ease of articulation. Yet, a speaker of German who suppresses the process of fronting of back vowels in alveolar context is not worse off than a speaker of English who applies this process. I.e., the suppression of the process does not pose an extra difficulty for the articulators of the speakers of German, otherwise they would not suppress it. It has been suggested that this process does not apply in languages which have front rounded vowels in their phoneme inventories, as e.g. German, French, Chinese (Oh 2002). However, in Cantonese, this process led to a neutralization of front rounded vowels in alveolar contexts (Flemming 2001), a fact that points to the application of this process despite the existence of front rounded vowels. In Standard Austrian German, as will be shown in 6.2, the application of this process is restricted to formal speech tasks in strong prosodic positions, a result which challenges the assumption that this process serves ease of articulation. Therefore, many ways for treating back rounded vowels in alveolar contexts can be observed cross-linguistically, and all of them are functional. It is, in any case, the phonology of a language or a language variety which decides which processes apply, and not the surrounding segments.

In the framework presented here, inertia, either from the side of the speaker or his or her speech organs, plays a marginal role. If these assumed antagonistic forces (clarity vs. ease) were decisive, no systematicity with respect to situational variability would be found, since any speaker would have different needs as regards his or her “economization” of speech gestures<sup>10</sup>. Speakers of one and the same social group are, however, quite consistent as concerns their speech behaviour, their application or suppression of processes, and their definition of the speech situation. Therefore, the concept of antagonistic forces is one-dimensional insofar as it does not consider different speaker – listener constellations, and, consequently, different interactional situations. Speech

---

<sup>10</sup> Since vocal tracts differ individually.

behaviour is, however, largely determined by the interactional situation, and the speakers are not interested in economizing their speech, but in carrying out a successful<sup>11</sup> interaction.

To ensure a successful interaction, speech behaviour is listener-oriented, and the phoneme is, therefore, to be conceptualized as a perceptual entity (Dressler 1979: 267). The speaker adjusts his or her articulatory configurations in such a way that the desired acoustic output is ensured. This implies that the speech chain is planned<sup>12</sup> (see also Perkell 1997, Donegan 2002), and that none of our output realizations are left to chance.

## 1.2. The phonological system

In Natural Phonology, the phonological system of a language consists of 1) prelexical processes, which constrain the number and combination of phonemes, 2) the phonemes, defined as sound intentions, and 3) postlexical phonological processes, which transfer the phoneme into the final phonetic output (Dressler 1984). The processes are

“mental substitutions which systematically but subconsciously adapt our phonological intentions to our phonetic capacities, and which, conversely, enable us to perceive in others’ speech the intentions underlying these superficial phonetic adaptations.” (Donegan & Stampe 1979: 126)

Processes are defined as responses to phonetic difficulties (Donegan & Stampe 1979: 136). However, only a minority of the processes existing in a language result from an irreconcilability of the phonological intention and the capability of the articulators. As concerns the prelexical processes, which are responsible for the phoneme inventory and the possible combinations of the phonemes in a given language, the selection of

---

<sup>11</sup> From a phonetic and phonological point of view, an interaction is successful, when the intended sequence of phonemes and their phonological processes are conveyed in a way that the listener is able to decode the perceived sequence and their processes. It does not include that, from a relationship-oriented point of view, the interaction is successful as well.

<sup>12</sup> According to Schütz (1962), any plan includes empty slots. These are responsible for psychological variables, e.g. attention or stress, which lie outside social and phonological intentions, but co-determine phonetic performance (Dressler & Moosmüller 1991: 136). The same holds for the consumption of drugs (Künzel et al. 1992).

phoneme combinations is language-specific. Therefore, what is difficult in one language need not be difficult in another one. For example, Tashlhiyt Berber exposes sequences of up to eight voiceless obstruents, as in “tftktstt” (you sprained it – fem, Ridouane 2002), a sequence which is not allowed in Germanic languages. Consequently, it is not justified to speak of processes as answers to phonetic difficulties, unless one defines a phonetic difficulty as language-specific. Language-specific phonetic difficulties do not exist; any child can learn any language as his or her mother tongue. Therefore, it is better to speak of language-specific preferences, which, once acquired, pose no difficulties on the speaker of the respective language. On the postlexical level, “difficulties” might occur when, superficially, combinations of sounds turn up which are excluded by prelexical processes in the given language, e.g. voiced obstruent + voiceless pause, leading to an assimilation of the voiced obstruent to the pause in German (better known as final devoicing), or voiceless plosive + voiced fricative, a sequence which is caused by the reduction of “that is all”<sup>13</sup> to “that’s all” in English, leading to the devoicing of /z/. Again, any language might respond differently to a sequence which is excluded on the prelexical level. A speaker of Albanian might rather voice the plosive than devoice the fricative in the sequence “that’s all”. The majority of postlexical processes are responses to the requirements of the interactional situation (e.g. the consecutive steps of nasal assimilation and consonant deletion in Standard Austrian German: /ha:bən/<sup>14</sup> → [ha:bm̩] → [habm̩] → [ham̩] → [ham] → [ɐm] depend on the degree of formality of the speech situation and the prosodic strength). Therefore, one should be very cautious in alleging ease of articulation or better pronounceability in a process, where, in most cases, postlexical processes serve figure-ground principles determined by the interactional situation. The speaker<sup>15</sup>, as long as he or she is

---

<sup>13</sup> Example from Donegan & Stampe 1979.

<sup>14</sup> The sequence bilabial plosive and alveolar nasal consonant is excluded in Standard Austrian German, therefore, the intermediate step \*[ha:b̩n̩], resulting from vowel deletion, is no observable surface realization.

<sup>15</sup> This holds, of course, only for speakers with no speech or hearing disorders and for children who have completed the acquisition of their language.

acquainted with the demands of a given interactional situation, acts, independent of whether a given sequence is “difficult” to pronounce or not, according to the demands of the interactional situation and according to the processes available in his or her language<sup>16</sup>.

It has been argued, in line with Natural Phonology, that the speech chain is planned. This plan contains all the fully specified information about the phonemic string plus the adaptations required for a successful interaction<sup>17</sup>. The adaptations performed are the postlexical processes which give the phoneme the ultimate shape. As soon as language acquisition is completed, the processes not suppressed for a given language or language variety start to become a habit. Due to this habit, speakers are not aware of many of the processes they apply or suppress. It is therefore argued, also in Natural Phonology, that some processes apply in a fairly or even fully automatized way. However, to apply processes automatically would imply that the speakers have no control over their plans and the execution of their plans. Having no or little control over one’s plans and the execution of one’s plans would never result in fluent speech<sup>18</sup>. Therefore, processes are habituated, and it is rather difficult to get rid of one’s habits. To apply and suppress different processes from the ones acquired in first language acquisition is trained – with speaker-specific success – in second language acquisition. A similar situation arises when speakers try to use a different variety from the one learned (e.g. dialect and standard). Speakers also become aware of their processes when speech production is hampered (e.g. by a bite-block) or when a sudden error occurs (see 6.4.).

Therefore, under normal conditions, the wish to convey a thought verbally, its planning, and the execution of the plan is intended by the speaker. The intention

---

<sup>16</sup> For example, sequencing vs. blending of articulatory gestures defined as conflicting.

<sup>17</sup> Van der Merwe (1997) distinguishes the linguistic-symbolic planning, including phonological planning, motor planning, and motor programming.

<sup>18</sup> As is the case with persons suffering from speech motor disorders, persons consuming drugs, or simply, persons whose attention is lowered (tiredness).

comprises both phonology and phonetics. Since both the phoneme and the final product of the phoneme are intended, the target has often been confused with the phoneme (see 6.1). The difficulty keeping phonetics and phonology apart has occupied phonological theory ever since Baudouin de Courtenay disunited phonetics into “Anthropophonik” and “Psychophonetik” (see Häusler 1976), and still tempts linguists to give up the distinction between phonetics and phonology (e.g. Port to appear). I will show in 6.3 that a phonological analysis is indispensable for a thorough understanding by which principles the final appearance of an utterance is lead (see also Moosmüller 2007a).

In order to avoid confusion, both a phonetic and a phonological analysis have to be performed. An output realization is, in any case, the result of the application of postlexical processes which more or less modify the phoneme. Where the output realization is identical with the phoneme<sup>19</sup>, no process is applied. Therefore, the application of no process is included in the plan (one could also speak of a zero process). Our knowledge about processes allows us to backtrace a phonetic output to its phoneme.

What then, is a phoneme? According to the most cited definition of Baudouin de Courtenay, it is

“eine einheitliche<sup>20</sup>, der phonetischen Welt angehörende Vorstellung, welche mittelst psychischer Verschmelzung der durch die Aussprache eines und desselben Lautes erhaltenen Eindrücke in der Seele entsteht = psychischer Äquivalent des Sprachlautes. Mit der einheitlichen Vorstellung des Phonems verknüpft sich (assoziiert sich) eine gewisse Summe einzelner anthropophonischer Vorstellungen, welche einerseits Articulations-Vorstellungen, andererseits aber akustische Vorstellungen, d. h. Vorstellungen gehörter oder im Gehörtwerden begriffener Resultate jener physiologischen Arbeiten, sind.“ (Baudouin de Courtenay 1895/1984: 65)

Whereas talking is impermanent, the phoneme is not. According to this definition, the phoneme is a mental imprint (“Eindrücke in der Seele”) which is altered according to our phonetic habits:

---

<sup>19</sup> For example in sustained vowel production tasks.

<sup>20</sup> Later on, Baudouin de Courtenay replaces “einheitlich” (uniform) with “ständig” (permanent) (Häusler 1976: 61).

“...zwingen uns unsere phonetischen Gewohnheiten, wie auch allgemein-menschliche Bedingungen phonetischer Verbindungen, die Aussprache der beabsichtigten Verbindungen ein wenig zu ändern...” (Baudouin de Courtenay 1895/1984: 75)

Why these alternations (that is how Baudouin de Courtenay termed processes) occur, stays a puzzle for him (“Diese Thatsache aber bleibt, was ihre Ursache betrifft, vorderhand rätselhaft” 1895/1984: 76).

The phoneme defined as sound intention is to be conceived differently from the intended target. The phoneme is not only an intended sound, it is also distinctive, and it works as a mental representative for the phonetic outputs associated with it via processes. The phoneme itself is the output of prelexical processes which are applied in a way that phonemes of a given language also perform a distinctive function. As will be argued in 3.1 and 4.4, an overcrowding of palatal vowels is averted in Standard Austrian German by the suppression of the prelexical process which merges constriction locations in the front area of the vocal tract. Consequently, Standard Austrian German distinguishes pre-palatal and palatal vowels in the front region. According to the requirements of the interactional situation and the prosodic position of the phoneme, in the case of Standard Austrian German vowels, postlexical processes change the degree of constriction, the length of constriction, the configuration of the lips, but they never change the location of the constriction<sup>21</sup>. The phonetic outputs might, as is well-known, overlap acoustically, because different articulatory configurations might lead to the same acoustic output. Each output can, however, be traced back unambiguously to its phoneme via our knowledge of processes. Therefore, it is vital for correct perception that we reanalyse the processes which determine a given phonetic output.

What does the phoneme look like? According to Dressler (1984), the phoneme is fully specified. It is usually taken for granted that a phoneme corresponds to an idealized pronunciation. This makes sense, since, besides the function of making up words (Donegan & Stampe 1979: 129), the phoneme should also be distinguishable from other phonemes, and the contrastive ability of phonemes decreases in weak

---

<sup>21</sup> Except in case of a sound change in progress, see 4.4.1.

prosodic positions or in interactional situations which demand a small degree of figure-ground contrast. Therefore, it is justifiable to assume that a phoneme corresponds to the pronunciation in a highly formal setting in stressed position (e.g. a reading task). Since, within this framework, the phonetic output is guided by the demands of the interactional situation, the phoneme is to be defined perceptually via the acoustic parameters with conclusions to be drawn on articulation, if possible. Since different articulatory configurations might result in identical or nearly identical acoustic outputs, it is the acoustic output that counts in an interactional situation, and not the articulatory configuration behind it.

Given the high variability dependant on the interactional situation together with the required figure-ground contrast, the phonological process is of high relevance for the analysis of a given phonetic output. Only the phonological process ensures an unambiguous trace back of a phonetic output to its underlying representation. Figure 1.1 gives an F1, F2, F3 plot over time of two examples of the vowel /e/, one in a strong prosodic position, the other one in a weak prosodic position, and one example of the vowel /ɛ/, in strong position. The F2 of the unstressed /e/ has more resemblance with the F2 of the stressed /ɛ/ than with the stressed /e/, the mental representative of the unstressed /e/. The F1 correctly goes with the stressed /e/, whilst the F3 shows hardly any differences between the three examples.

Models which deny phonological processes (e.g. Keating 1990) face a confusing situation: does this unstressed vowel belong to /e/ (correspondence of F1) or to /ɛ/ (correspondence of F2)? However, both speakers and listeners know that in the unstressed position, the constriction degree is widened and the constriction length is shortened for /e/, leading to a lowering of F2, whereas the degree of lip opening is not changed. Therefore, due to the phonological process, both speaker and listener will unambiguously trace the unstressed vowel back to its phoneme /e/.

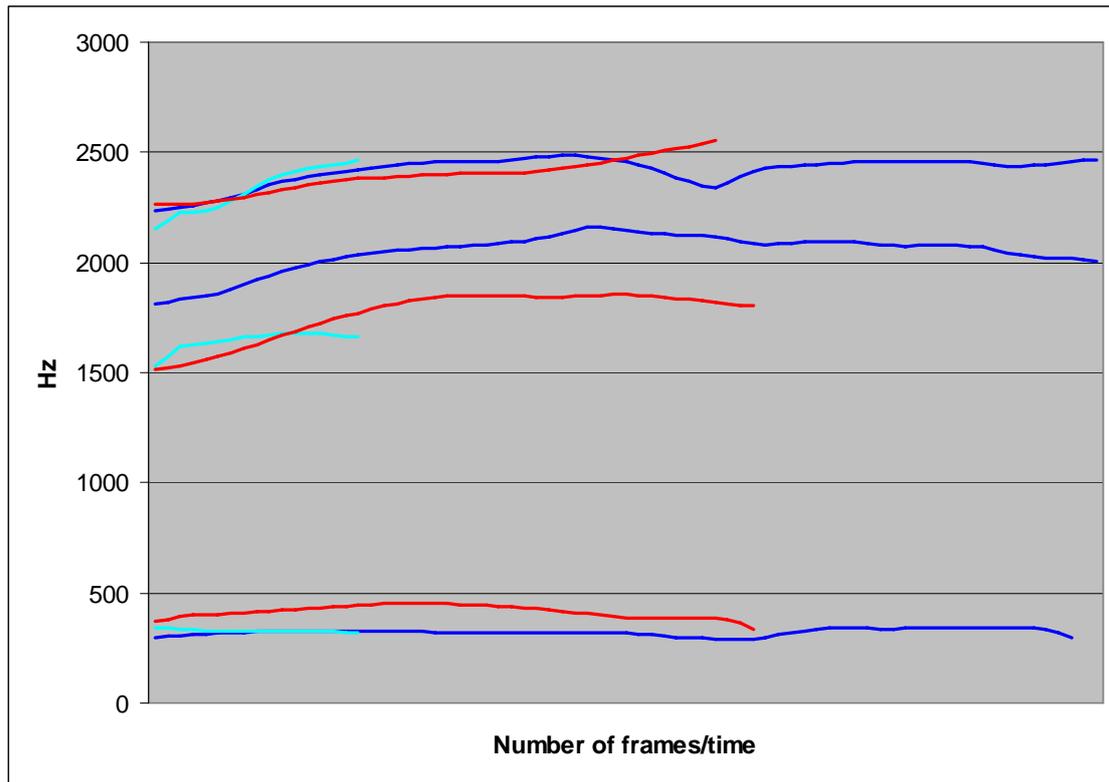


Figure 1.1: Spectral change over time of F1, F2, and F3 of the vowel /e/ from the token “Beet” (patch), blue lines, the vowel /e/ of the token “Bedeutung” (importance), turquoise lines, and the vowel /e/ of the token “besser” (better), red lines. Speaker sp127, sentence reading task. No time alignment has been performed.

### 1.3 Standard Austrian German

Standard Austrian German is, in the same way as any standard language, both a regional and social variety, which is, contrary to the other varieties of a given speech community, accepted by the majority of the members of the speech community to function as the standard language. Standard languages are the result of political unification processes (Barbour & Stevens 1990) and depend to a larger extent on political borders than on dialect borders (Reiffenstein 1983). Therefore, standard language’s function is to separate a given language from other languages (across borders) and to unify the varieties of a given language within borders (Dittmar 1997)<sup>22</sup>.

<sup>22</sup> Dittmar (1997) also names prestige and reference for correctness as functions of a standard language.

This unifying function is often misconceived as a production criterion, i.e. the standard language is assumed to have no regional characteristics in order to be able to fulfill its unifying function. Apart from the fact that within such a framework the standard language is allowed to show social characteristics, since standard languages are usually spoken by the elites of a given speech community, supraregionality is in most cases unrealistic, given the regional diversities in most political entities. Therefore, in historically grown standard languages, the standard language is either the result of a counterbalance of the political and cultural elites, as e.g. in Dutch (Haar 2001), or the variety of those holding the political power, see e.g. Queen's English (Schröder 2001). Or, the standard language is completely left to the social elites, irrespective of their regional background, as is the case for Dansk (Zynt-Dyhr 2001). Modern Standard Albanian, a young standard, is primarily based on the Tosk variety favoured by the political powers of the time. However, counterbalancing strategies of the social and cultural elites can be observed as well (Moosmüller & Granser 2006). Therefore, a standard language is neither supraregional nor suprasocial in production, it is however, supraregional in acceptance.

Dittmar (1997) names six criteria<sup>23</sup> necessary for defining a variety as a standard; of these, Standard Austrian German lacks the first, the written codification. Especially Coulmas (2000) emphasises the necessity of a codified reference variety for the development of a standard language. Nevertheless, the great majority of Austrians hold the opinion that an independent standard exists for Austria, and they are, moreover, quite consistent in what this standard variety looks like (Moosmüller 1991). Assessment tests performed by Moosmüller (1991) reveal that the phonology of Standard Austrian German is based on the Middle Bavarian varieties and is spoken by the educated people with a social background of the upper and middle social classes. Regionally, the

---

<sup>23</sup> The six criteria are: written codification, supraregional spread and acceptance, usage in formal contexts, difference to everyday usage, sanctions in case somebody does not know the standard language, and high prestige.

standard is located in the large cities, i.e. Vienna, Linz, and Salzburg. These varieties are accepted as supraregional. The samples of the educated people of the South Bavarian varieties are recognized and accepted as “supraregional” as long as they lack South-Bavarian characteristics. Otherwise, they are attributed a high social status, but without supraregional acceptance. These results hold for people of Middle-Bavarian and South-Bavarian varieties and for all social classes<sup>24</sup>.

Therefore, Standard Austrian German lacks processes attributed to Standard German (separating function), e.g. vowel deletion between nasal consonants, it lacks South-Bavarian characteristics, e.g. suppression of r-vocalization or diphthongization of stressed, constricted vowels, and it lacks salient dialectal input switch rules<sup>25</sup>. Middle-Bavarian dialectal processes are allowed in the prosodically weakest positions, e.g. l-vocalization in function words.

---

<sup>24</sup> For a detailed discussion see Moosmüller (1991).

<sup>25</sup> See Dressler & Wodak (1982), Moosmüller (1987, 1991) for the interaction model of dialect and standard and for examples of input-switch-rules.

## 2. Methods and Data

### 2.1. Speakers and Material

Three female and three male speakers of Standard Austrian German were asked to act as speakers. According to the results presented in Moosmüller (1991) on Standard Austrian German, the speakers were chosen according to the following criteria:

- Variety: middle Bavarian
- Region:
  - ◆ At least one parent was brought up in Vienna<sup>26</sup>
  - ◆ The speakers themselves were brought up in Vienna
- Educational background:
  - ◆ At least one parent has an academic education
  - ◆ The speakers themselves have either an academic education or a school-leaving certificate<sup>27</sup>.

According to these criteria, the following speakers were chosen:

- Speaker sp082<sup>28</sup>: University professor, female, 40 years<sup>29</sup>
- Speaker sp129: University professor, female, 49 years
- Speaker sp180: Statistician, female, 26 years
- Speaker sp012: Researcher, male, 36 years
- Speaker sp126: Architect, male, 46 years

---

<sup>26</sup> Vienna was chosen for practical reasons, and not because the author holds the opinion that Standard Austrian German is solely spoken in Vienna (see 1.3 and Moosmüller 1991).

<sup>27</sup> Depending on the age of the speaker.

<sup>28</sup> The database of Standard Austrian German is sorted chronologically: speakers with a lower number have already participated in previous projects and were willing to give their time once more for this venture.

<sup>29</sup> At the time of recording (2002/2003).

- Speaker sp127: Student, male, 20 years

Open interviews were carried out with the speakers in a sound proofed room (labeled ‘spontaneous speech’). The interview contains data about social and educational background, the region the speakers and their parents were brought up, and a conversation on topics which arose in the course of talking. The interview comprises approximately 20 minutes of speech. After that, the speakers were asked to read a list of 72 sentences, twice (see Appendix).

It is generally agreed that speech behaviour changes with respect to the speech situation. However, in the current investigation, the speech situation stayed the same, i.e. speakers were recorded in the same sound proofed room. Therefore, it was decided not to refer to different “speech situations”, but rather to different “speaking tasks” in connection with the current data. If appropriate, when general statements or remarks are made, the term “speech situation” is used as well.

In a separate session, two speakers (sp012 and sp180) were also asked to read a list of bisyllabic logatomes in a carrier sentence (“Er hat sich PVPe<sup>30</sup> genannt.” – “He called himself bebe.”). The first (stressed) syllable of each logatome changed with respect to vowel and consonantal environment. 14 phonological vowels were assumed: /i, ɪ, y, ʏ, e, ɛ, ø, œ, u, ʊ, o, ɔ, a/. Plosives were chosen as consonantal environment: /b, p, d, t, g, k/. For each vowel pair ([+constricted] and [–constricted]), 12 different consonantal environments were produced, according to the following pattern:

*“Er hat sich bebe genannt”<sup>31</sup>*

*“Er hat sich bepe genannt“*

*“Er hat sich bebbe genannt”*

*“Er hat sich beppe genannt”*

This list, presented on index cards in randomised order, was read twice.

<sup>30</sup> P = plosive, V = vowel.

<sup>31</sup> The first two items (“bebe” and “bepe”) are supposed to yield the [+constricted] vowel /e/, the second two items (“bebbe” and “beppe”) the [–constricted] vowel /ɛ/.

The speech material thus obtained was digitised at 22.050 Hz, 16 Bit with the workstation STx (<http://www.kfs.oeaw.ac.at>).

## 2.2. Measurements

In total, approximately 11.000 vowels were analysed. The vowels were segmented manually. In the case of preceding voiceless segments, the first positive zero crossing or the end of aperiodicity was determined as the start of the vowel. The last full period similar to the preceding ones, or the start of aperiodicity was determined as the end of the vowel. Figure 2.1 gives an example for segmentation of a vowel in a fricative – plosive context. The cursors in the spectrogram window are positioned at the start and the end of the vowel. The waveform zoom window (left upper panel) at position 0<sup>32</sup> ensures the exact positioning at the first positive zero crossing, referring to the left cursor position. The segmentation served as a duration measurement as well.

Formant frequency candidates were extracted by means of a Linear Prediction Coding (LPC) algorithm such as published by Markel & Gray (1976). At a sampling frequency of 22.050 Hz, a 46 ms long gliding Hanning window was applied with an overlap of 95%, using 26 coefficients and providing sufficient measurement points for fast formant transitions and short signal segments. Fundamental frequency measurements were performed by means of an autocorrelation method (SIFT: Simplified Inverse Filter Tracking), synchronized with the formant frequency analysis.

---

<sup>32</sup> Not to be confused with the cursors (crosses) in the waveform zoom window, which are positioned to measure the duration of the period, with the values displayed in the left column at the bottom.

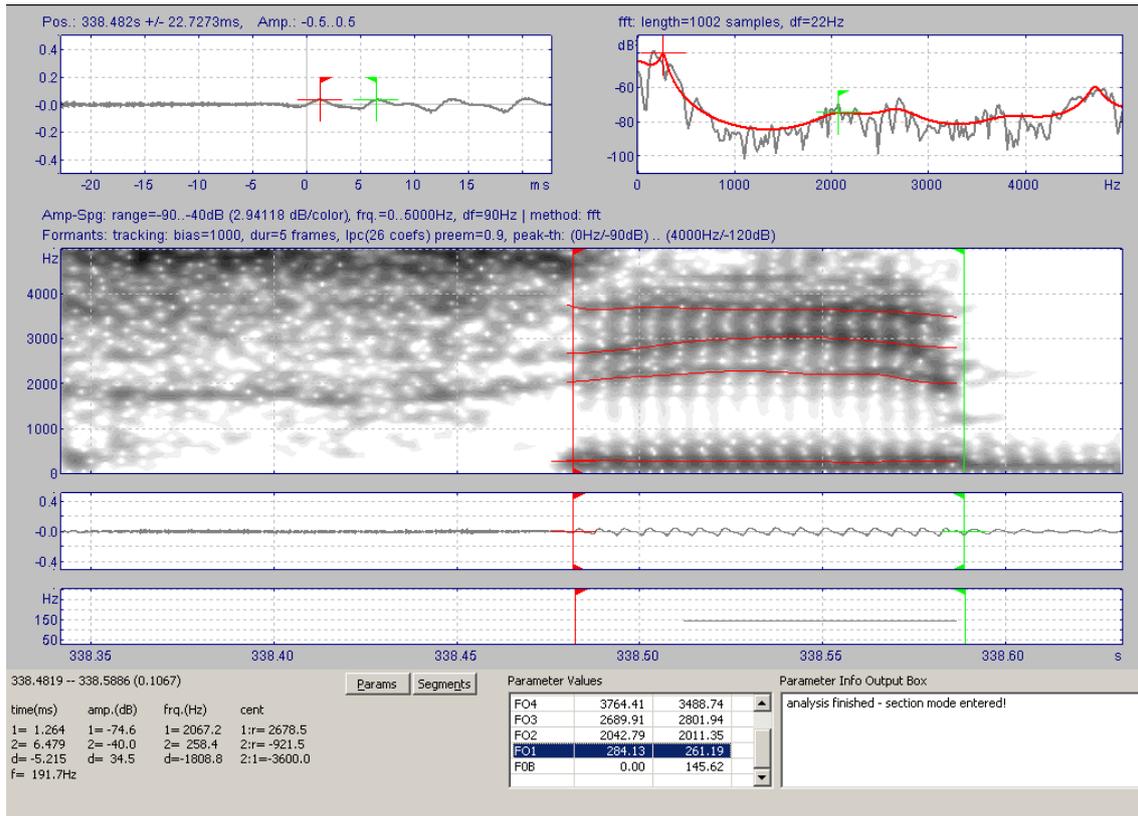


Figure 2.1: Spectrogram of the item “siebe” (sieves: PL). Speaker sp127, sentence reading task. Bottom panel: fundamental frequency. Next panel from bottom: waveform window. Third panel from bottom: spectrogram window. Left upper panel: waveform zoom window. Right upper panel: amplitude spectrum window.

Where preceding sonorant consonants (nasals and liquids) occur, segment borders were determined by simultaneous spectrogram, waveform and amplitude inspection. Vowel – liquid sequences were not analyzed, because, in the case of a subsequent lateral, the transition from vowel to the lateral is usually long and the change takes place monotonously; therefore, no meaningful segment border can be found. Additionally, speaker specific handling of the sequence would render inconsistent segmentation and, consequently, incomparable data. In the sequence /Vr/, the trill is vocalized under certain conditions (see Moosmüller 1991) in German, leading to a diphthongal movement. In the same way as in the sequence /VI/, no meaningful segment border can be found. Moreover, in an unstressed position, this diphthong might be monophthongized. Figure 2.2 captures two items; the sequence “stern” from the item

“Sternzeichen” (star sign) and the item “dieser” from the sequence “dieser Kopf” (this head).

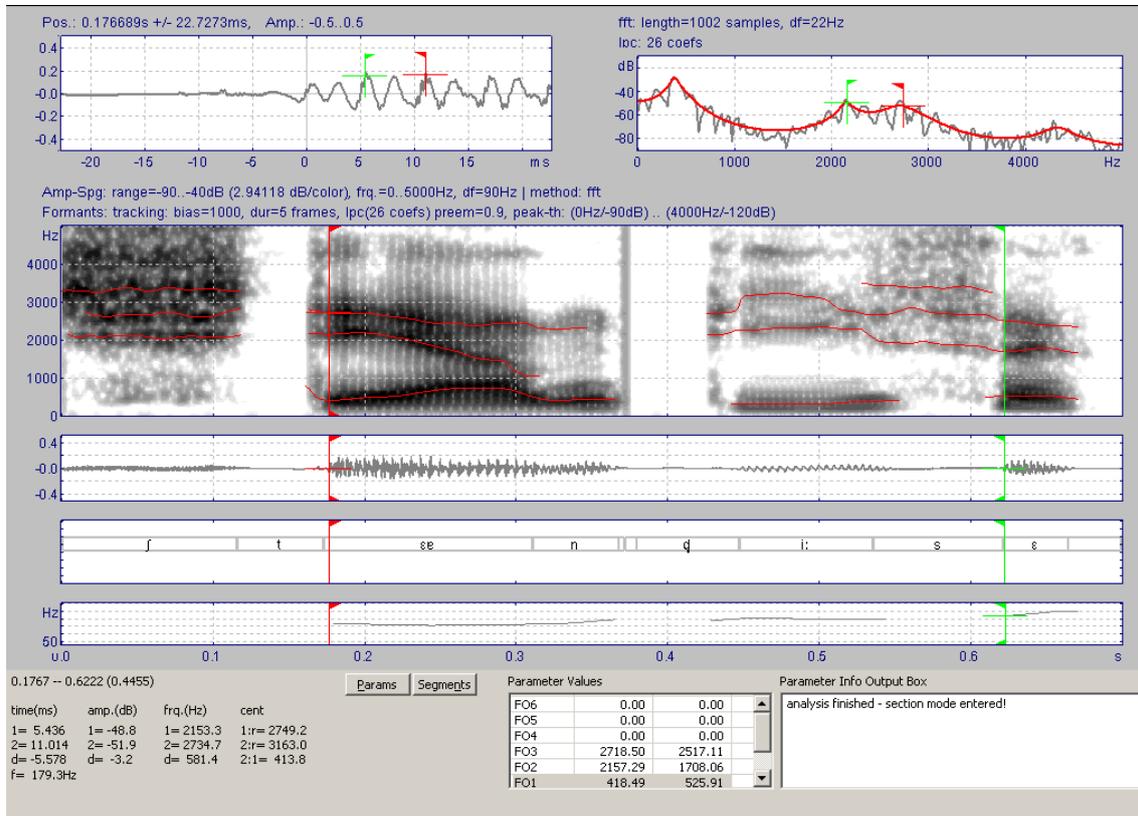


Figure 2.2: Spectrogram of the sequence “stern” from the item “Sternzeichen” (star sign) and the item “dieser” from the sequence “dieser Kopf” (this head). Speaker sp129, sentence reading task. Bottom panel: fundamental frequency. Next panel from bottom: phonetic transcription. Third panel from bottom: waveform window. Fourth panel from bottom: spectrogram window. Left upper panel: waveform zoom window. Right upper panel: amplitude spectrum window.

Cursors mark the onset of the phonetic realization of the sequence “er” respectively<sup>33</sup>. The waveform zoom window and the amplitude spectrum window refer to the left cursor in the spectrogram window. In both cases, the trill is vocalized. The sequence “er” from the item “Sternzeichen” is in a stressed position and is realized as a diphthong; this is clearly indicated by the movement of the formants. The second sequence “er” from “dieser” is in an unstressed position; therefore, the diphthong is

<sup>33</sup> The thick grey bar in the spectrogram window marks the border between “stern” and “dieser”.

monophthongized, realized in this case as /ɛ/<sup>34</sup>. Especially in the weak prosodic position, Standard Austrian German shows a strong tendency towards monophthongizing both phonological diphthongs and diphthongs derived from phonological processes (for an analysis of Austrian German diphthongs, see Moosmüller 1997b, c, Vollmann & Moosmüller 1999, Moosmüller & Vollmann 2000, 2002).

The sequence lateral – vowel might cause some problems as well (see Figure 2.3).

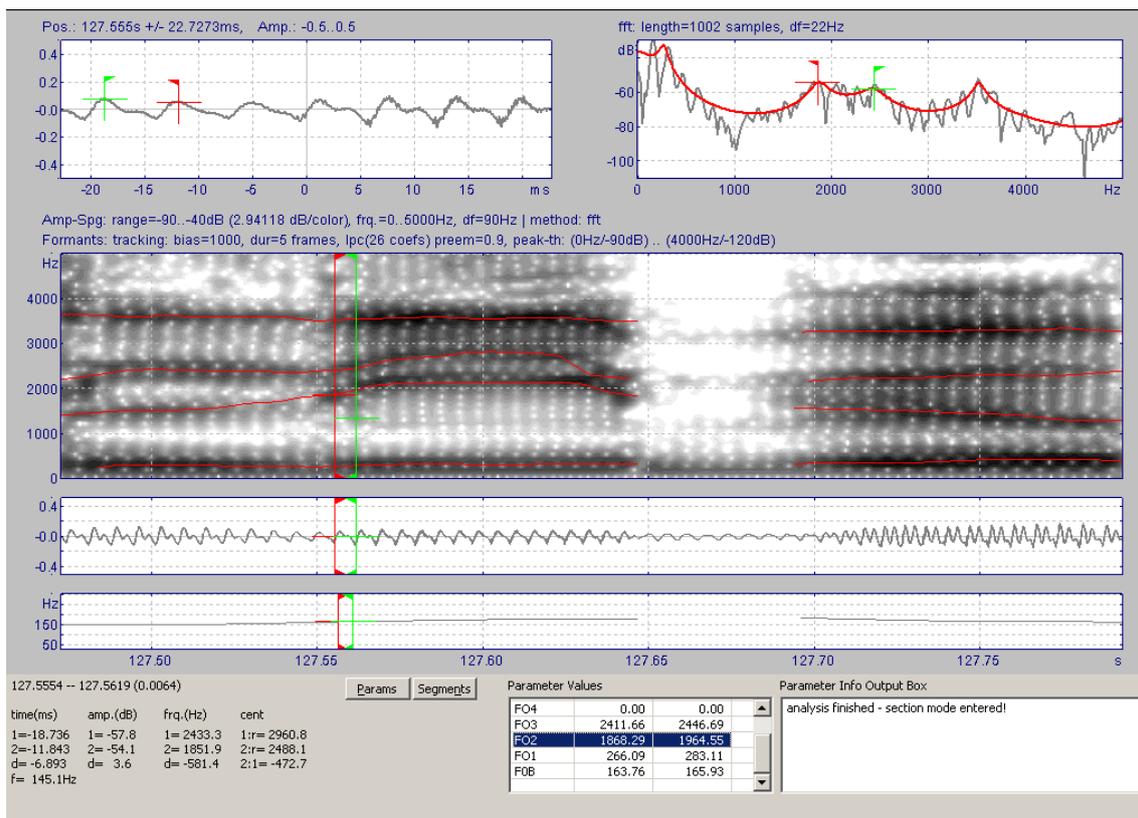


Figure 2.3: Spectrogram of the item “liebe” (dear: ADJ). Speaker sp127, sentence reading task. Bottom panel: fundamental frequency. Next panel from bottom: waveform window. Third panel from bottom: spectrogram window. Left upper panel: waveform zoom window. Right upper panel: amplitude spectrum window.

Usually, the segment border between lateral and vowel is detectable by a higher amplitude for the vowel and, for some speakers, by a plosion at the point where the tongue is adjusted for the vowel. However, sometimes, the transition is quite monotonous; in these cases, an inaccuracy of  $\pm$  one period is accepted. Figure 2.3 gives

<sup>34</sup> There are many possibilities for realizing the phonetic output of the monophthongization process, even within one speaker’s output (see Moosmüller 1997a).

an example of a monotonous change of formant frequencies. The period in question is put between the two cursors in the spectrogram window. The position “0” in the waveform zoom window refers to the left cursor in the spectrogram. The period in question does not enter the analysis. Vowel – semivowel, semivowel – vowel, and vowel – vowel<sup>35</sup> sequences were not analyzed for the same reason as put forward for vowel – liquid sequences.

From the vowels selected for analysis, F1, F2, F3, F0, and duration measurements have been performed. Depending on the duration of the vowel, the measurement procedure described above rendered 20 – 150 measurements per vowel. This data was exported for further statistical analysis. Except where otherwise mentioned, the formant frequency contour of the whole vowel was analysed (i.e., transitions enter into the mean values). This method was chosen because vowels, especially when short, often expose no steady state portion. Therefore, an analysis relying solely on the steady state portion or the vowel midpoint would not produce reliable results. Figure 2.4 gives an example of a vowel with no steady state portion for F1 and F2: formants monotonously move from the bilabial plosive to the velar plosive (see also Chapter 5). The vowel in question is displayed between the two cursors in the spectrogram window. The waveform zoom window and the amplitude spectrum window refer to the right cursor position in the spectrogram.

---

<sup>35</sup> For an analysis of quasi-homorganic vowel – vowel sequences in Austrian German see Moosmüller (1999).

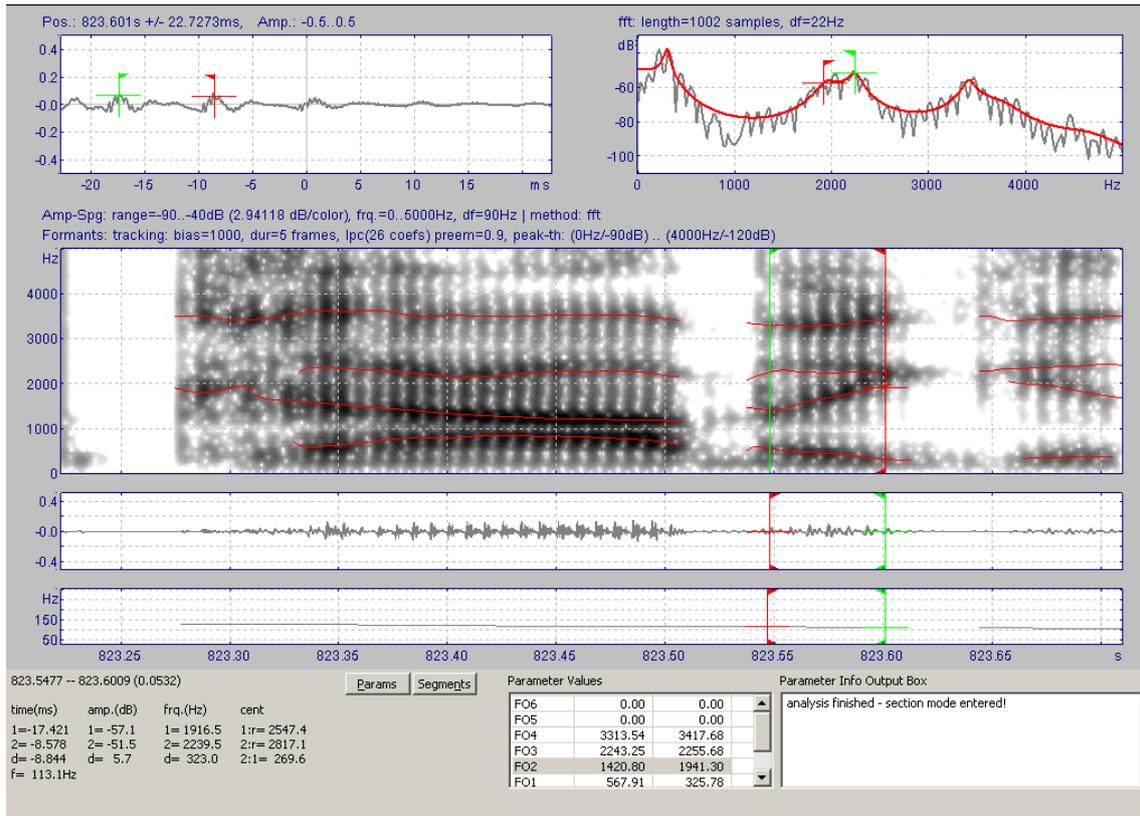


Figure 2.4: Spectrogram of the sequence “Grabe ge-” from the sequence “Grabe getragen” (fig., die out: PT). Speaker sp127 sentence reading task. Bottom panel: fundamental frequency. Next panel from bottom: waveform window. Third panel from bottom: spectrogram window. Left upper panel: waveform zoom window. Right upper panel: amplitude spectrum window.

If necessary, formant frequency traces were corrected or aligned to the correct order. Figure 2.5 presents the calculated formant frequency movement and alignment in an uncorrected form. As concerns the vowel /i/ (between the cursors), the respective formants have been identified correctly. However, the movement of F3 is interrupted for quite a substantial part of the total duration of the vowel. STx provides several possibilities for editing and correcting.

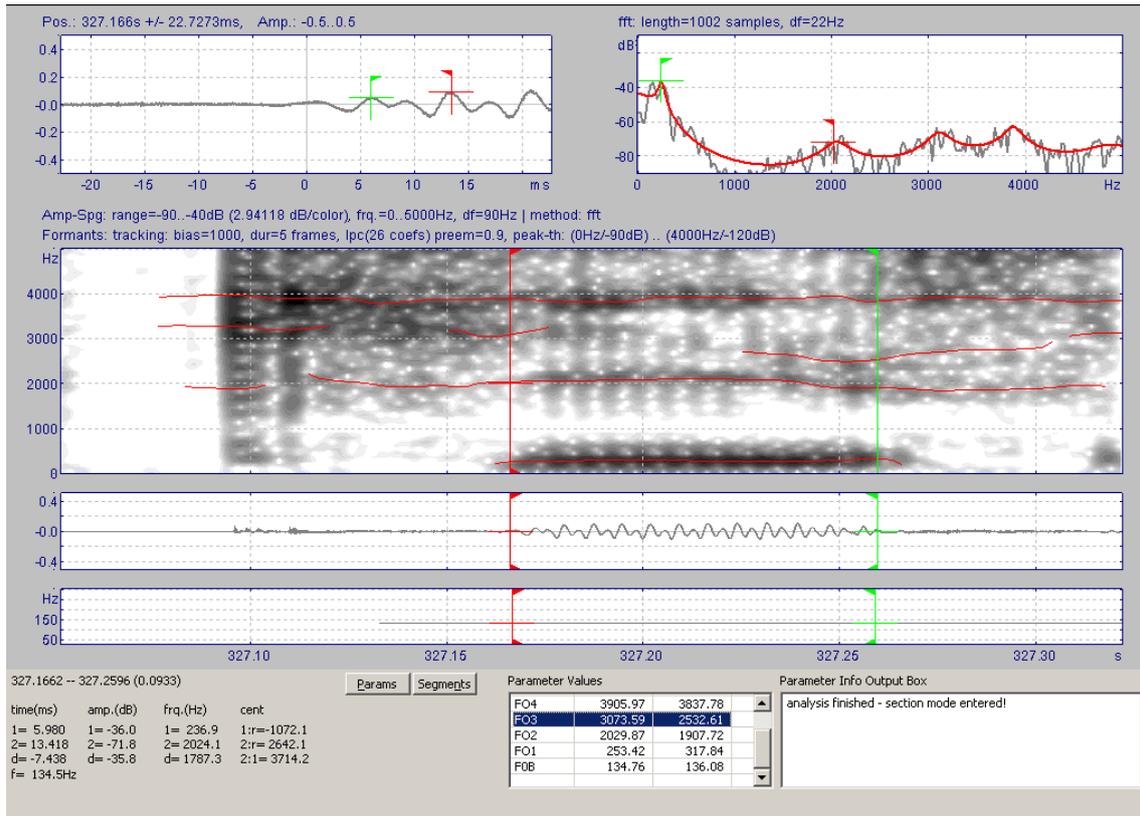


Figure 2.5: Spectrogram of the item “Kies” (gravel) where formants are not corrected. Speaker sp012, sentence reading task. Bottom panel: fundamental frequency. Next panel from bottom: waveform window. Third panel from bottom: spectrogram window. Left upper panel: waveform zoom window. Right upper panel: amplitude spectrum window.

In the present case, a value for F3 at 327.18 s was estimated and entered. Subsequently, the data points were joined. The result of the correction is presented in Figure 2.6:

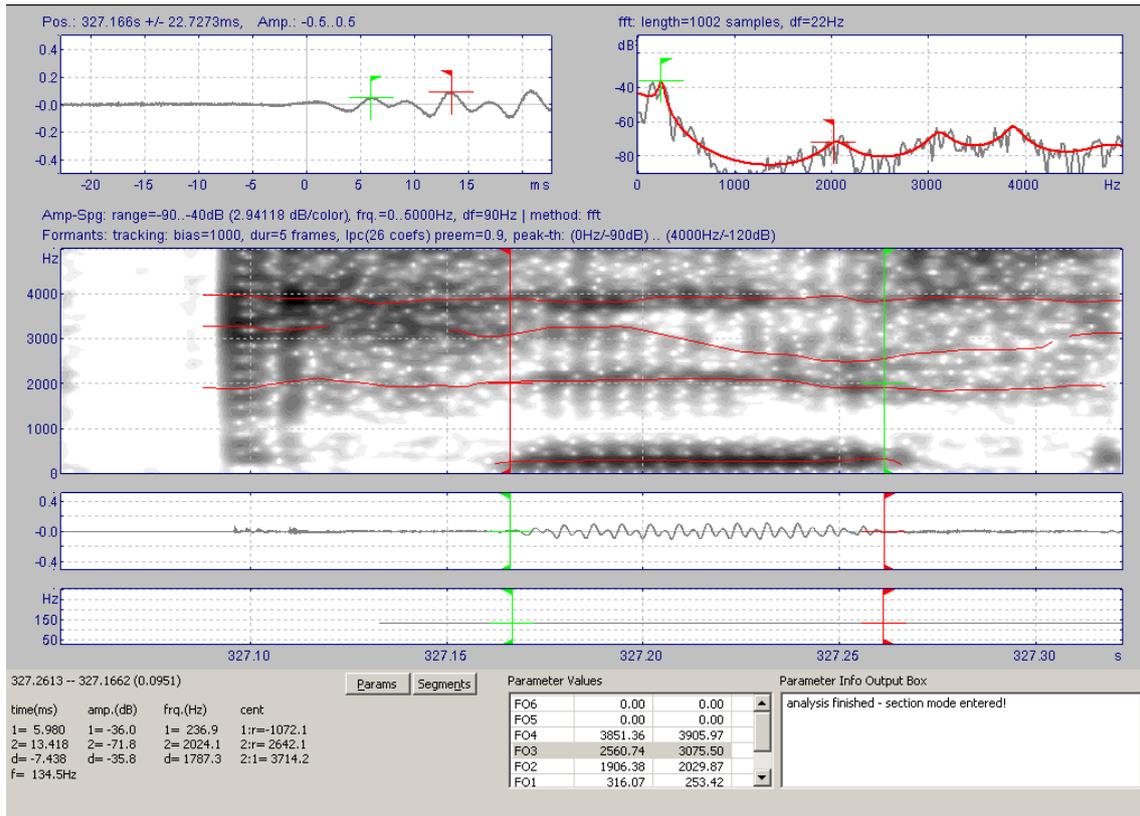


Figure 2.6: Spectrogram of the item “Kies” (gravel) with F3 corrected. Speaker sp012, sentence reading task. Bottom panel: fundamental frequency. Next panel from bottom: waveform window. Third panel from bottom: spectrogram window. Left upper panel: waveform zoom window. Right upper panel: amplitude spectrum window.

It is well known that measurement problems arise when two formants start to approach each other. Especially some [+ constricted] vowels might be affected by convergence: F1 and F2 for /u and o/, F2 and F3 of /y and e/. In this study, the respective formant frequency was estimated where possible and, if no estimation was possible, the vowel was discarded from analysis. In Figure 2.7, the left cursor is placed at the vowel midpoint of the vowel /o/ of the item “Sohn” (son). The waveform zoom window and the amplitude spectrum window refer to the left cursor position. It can be seen from the amplitude spectrum window that F1 and F2 converge; therefore, this vowel was not included in the analysis.

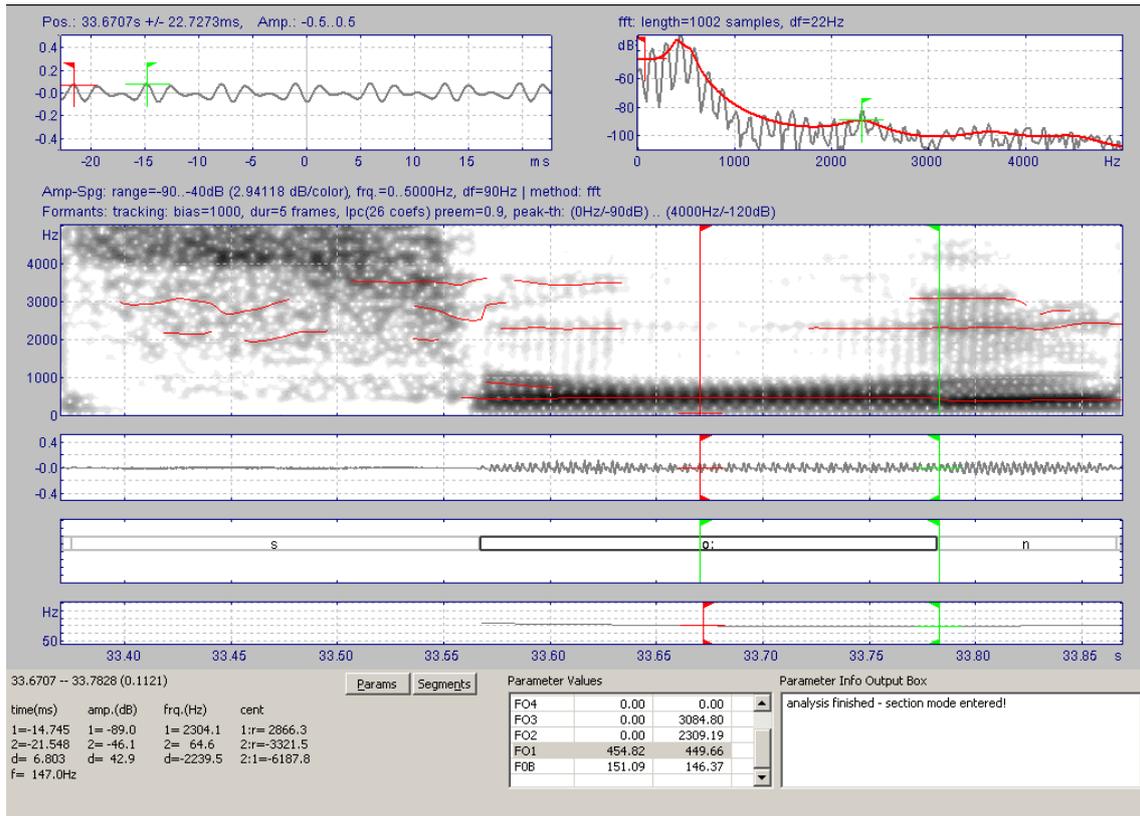


Figure 2.7: Spectrogram of the item “Sohn” (son). Speaker sp129, sentence reading task. Bottom panel: fundamental frequency. Next panel from bottom: phonetic transcription. Third panel from bottom: waveform window. Fourth panel from bottom: spectrogram window. Left upper panel: waveform zoom window. Right upper panel: amplitude spectrum window.

### 3. Interpretation of formant measurements

In the same way as vocal tract shapes differ individually, articulatory configurations expose individual differences as well. Many studies on articulation report speaker-specific articulatory settings for either the same phoneme or sequence of phonemes (see e.g. Kuehn & Moll 1976, Perkell 1997, Perkell et al. 2002, Tabain et al. 2004, Pouplier et al. 2004, McGowan 2004, Brunner et al. 2005, Perkell et al. 2006). However, due to the non-linear relation between articulatory configurations and the acoustic consequences of these configurations, these differences need not affect the acoustic output. A study of Fant's nomograms reveal that different articulatory configurations might render the same formant frequency values. A study of Steven's quantal theory exposes that, on the horizontal dimension from front to back, the extent to which formant frequency values change depends on where the constriction is located. Therefore, it is necessary to discuss what the acoustic data are able to tell us.

#### 3.1. The traditional F1/F2 representation

To date, vowels are frequently represented in a two-dimensional F1/F2 chart (for German see e.g. Jørgensen 1969, Kohler 1998, for Austrian German Iivonen 1987b) and interpreted with respect to the first two formants. Figure 3.1 gives such a representation for speaker sp012 of Standard Austrian German:

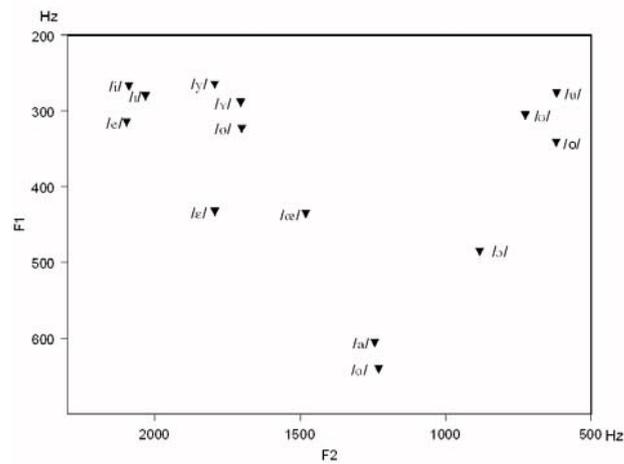


Figure 3.1: F1/F2 plot of the vowels of Standard Austrian German spoken by speaker sp012, logatome reading task.

This chart looks quite symmetrical, it suggests, however, a neutralization of /i/ and /e/, of /y/ and /ø/, of /u/ and /o/ and of /a/ and /ɶ/ on the F2 scale, and nearly a neutralization of /i/ and /ɪ/ on the F1 scale. A similar observation can be made for speaker sp180, as becomes evident from Figure 3.2:

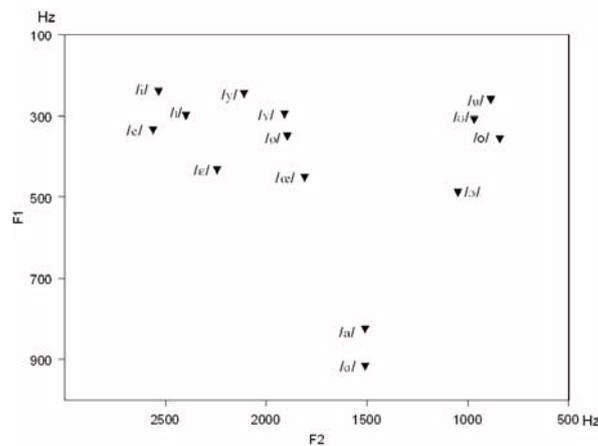


Figure 3.2: F1/F2 plot of the vowels of Standard Austrian German spoken by speaker sp180, logatome reading task.

Moreover, in the upper part of the vowel space, the vowels are located quite closely together on the F1 scale. Similar results have been obtained for Danish (Ejstrup &

Hansen 2003), German<sup>36</sup> (Kohler 1998), and Fering (Bohn 2004). Ejstrup & Hansen (2003) propose a sound change in progress, and Bohn (2004), in comparing the vowel spaces of Southern British English, North German, Danish and Fering, asks

“whether the uneven distribution of vowels in the vowel space is a general feature of languages with large vowel inventories, or whether we are dealing with a *Sprachbund* phenomenon of Fering, North German and Danish, which are spoken in neighboring and overlapping geographical areas.” (Bohn 2004: 165)

It will be argued that the traditional F1/F2 space insufficiently represents vowels<sup>37</sup>, especially if we have to deal with large vowel systems. Since the frequently cited investigation of American English vowels performed by Peterson & Barney (1952), many studies on vowels conclude that the first two formants are the most important acoustic parameters for vowel quality distinction (Harrington & Cassidy 1999:60) and, consequently, sufficient to represent the vowels of a given language or language variety.

The convention of representing the formant data of vowels in an F1/F2 plot goes back to Joos (1948) who demonstrated the relationship between the F1/F2 representation and the concept of the vowel quadrilateral, as well as showing that F1 and F2 are negatively correlated with vowel height and backness respectively.

“A brief glance at this diagram shows that the correlation between articulation and vowel color is (at this stage in the investigation) astonishingly simple. Although the vowel samples have here been placed on the chart strictly according to acoustic measurements (made from a phonograph record!), the diagram is practically identical with the classical ‘tongue position’ chart. ... Of course the scales of these diagrams were deliberately set up so as to enhance the resemblance of the acoustic chart to the tongue-position chart. For the directions in which the two scales run—toward the left and downward, contrary to usual graphical practice—the reason was that this puts [i] at top left, [u] at top right, and [a] at the bottom, to agree with the usage of the International Phonetic Association, and for this no apology is needed.” (Joos 1948: 53)

The concept of the vowel quadrilateral in turn goes back to Hellwag (1781, cited in Stratka 1978), who represented the vowels in an isosceles triangle according to the

<sup>36</sup> The speakers of the Kiel Corpus come from the north-western part of Germany. The variety spoken has been termed “Northern Standard German” by Wesener (2001).

<sup>37</sup> This is nothing new and has e.g. been pointed out by Pols et al. (1969). Jørgensen (1969) plotted the vowels in an F1/F2’ (effective F2 which takes F3 into account) space, but this did not enhance the discrimination of front rounded and unrounded vowels, therefore he returned to the traditional F1/F2 space. See also Pfitzinger 2005 for a critique.

placement of the highest point of the tongue. This vowel triangle has undergone many changes and took many shapes<sup>38</sup>, till it ultimately resulted in a quadrilateral with a perpendicular line for the specification of the back vowels and a line with an obtuse angle for the front vowels (Handbook IPA 1999). It did not escape Joos' notice that the IPA vowel quadrilateral does not exactly reflect articulatory tongue positions drawn from X-ray data and he drew the conclusion that the IPA representation is actually an acoustic one:

“Comparison of Figs. 26, 27, and 28 leads to a very interesting tentative conclusion. It appears that the classical or IPA tongue-position quadrilateral rather more closely resembles the acoustic vowel quadrilateral than it resembles the Carmody X-ray tongue-position quadrilateral. ... It should be noted that nowadays the IPA chart officially has an articulatory shape but an acoustic shape.” (Joos 1948: 54)

Joos (1948) has especially pointed out the discrepancy in the spacing of [i], [e], and [ɛ] between the articulatory and the IPA chart. The focus later turned to the representation of back vowels in the diverse vowel charts:

“Mais aucune de ces figures ne correspond, dans sa partie postérieure, à la réalité.” (Stratka 1978: 440).

Regardless of whether one takes the highest point of the tongue or the constriction location as a reference point, /o/ is articulated further back than /u/ (see e.g. Stratka 1978, Wood 1979, Pétursson 1992<sup>39</sup>, Ladefoged & Maddieson 1996, Fant 2001). From the point of view of constriction location, /u/ is the only vowel that can be described as velar; the rest of the back vowels have their constriction in the pharyngeal region. The IPA Handbook circumnavigates this problem by defining the intermediate vowels auditorily:

“Specifically, two fully front vowels [e] and [ɛ] are defined between [i] and [a] so that the differences between each vowel and the next in the series are auditorily equal; and similarly, two fully back vowels [ɔ] and [o] are defined to give equidistant steps between [ɑ] and [u]. The use of auditory spacing in the definition of these vowels means vowel description is not

---

<sup>38</sup> Though discarded since long in phonetics, Hellwag's vowel triangle persists in other scientific areas, as becomes obvious from the following example: “According to the position of the tongue in the oral cavity vowels are divided into high (i, u), central (e, o) and low (a) – the so-called Hellwag triangle.” (Jindra et al. 2002: 91).

<sup>39</sup> Pétursson proposes a vowel trapezoid which meets the fact that [ɑ] is located further back than [o], which again is further back than [u] (Pétursson 1992: 45).

based purely on articulation, and is one reason why the vowel quadrilateral must be regarded as an abstraction and not a direct mapping of tongue position.” (Handbook IPA 1999: 11f)

A half articulatorily and half auditorily based representation causes a confusing situation; consequently, Ladefoged & Maddieson (1996) define the classical vowel quadrilateral auditorily, with reference to acoustic mapping (F1 plotted against F2-F1):

“The acoustic representation corresponds more closely to the auditory phonetic description in terms of height and backness than the articulatory plots in figures 9.2-9.4.....Recognition that the placement of vowels on an auditory chart such as the one in figure 9.1 [the traditional vowel quadrilateral] is supported more readily by acoustic than by articulatory measurements does not mean that articulatory scales can be discarded in the phonetic description of vowels.” (Ladefoged & Maddieson 1996: 285)

Lindblom (1986) emphasizes three facets in defining the vowel space for a language: the articulatory stage, the acoustic stage, and the auditory stage. In the theory of adaptive dispersion (TAD; Liljencrants & Lindblom 1972, Lindblom 1986, 2003, Diehl & Lindblom 2004), the focus is directed towards the acoustic-auditory mapping. This theory is based on the concept that “distinct meanings must sound different” (Lindblom 2003) and that consequently, the dispersion of vowels in the “available phonetic space” (Diehl & Lindblom 2004) meets the principle of maximal contrast (Liljencrants & Lindblom 1972, Lindblom 1986). Liljencrants & Lindblom (1972) tested this hypothesis by first defining the shape of a universal vowel space. Secondly, the perceptual contrast between any two vowels was measured. To measure the perceptual contrast, euclidean distances between the formant frequencies in Mel units were calculated. For reasons of simplification, formant frequencies were restricted to two dimensions,  $M_1$  and  $M_2'$ <sup>40</sup>. Thirdly, on the basis of intervowel distances, optimal vowel inventories were created. For inventories of up to six vowels, the predicted systems were identical to those of the preferred vowel inventories. For larger inventories, too many high vowels were predicted. Modifications of this early model (Lindblom 1986,

---

<sup>40</sup>  $M_2'$  =  $F_2$  corrected to reflect the spectral contributions of  $F_3$ , in Mel units.

Diehl et al. 2003) yielded better results, nevertheless, the results remain unsatisfactorily, especially for larger vowel systems<sup>41</sup>.

In the Liljencrants & Lindblom (1972) model, a two-dimensional representation is favoured, a three-dimensional auditory-perceptual space has been proposed by Miller (1989) who transformed the measurements of F0, F1, F2, and F3 into log frequency ratios and plotted them as points in a three dimensional space. The vowels of each category were enclosed in three dimensional target zones. Miller (1989) analysed the nine monophthongal vowels of American English and could, by this method, create nine non-overlapping target zones which account for 93% of the data.

This method has also been applied to German and Greek vowels (Jongman et al. 1989). The zones created could differentiate the five vowels of Modern Greek with 100% accuracy, and the fourteen vowels of German with 94% accuracy. Jongman et al. (1989) conclude:

“In general, it would seem advantageous for a given language to have vowels that are maximally distinct acoustically (see, for example, Liljencrants & Lindblom 1972; Stevens 1972; Lindblom 1986) for reasons of communicative efficiency. Greek provides an example with its five vowels being quite far from each other in APS. Interestingly, five-vowel inventories similar to that of Greek are much more frequent than any other type of vowel inventory....The vowel spaces of German and American English are much more dense. It seems that the larger the vowel inventory, the more peripheral the location of the extreme vowels (...), relative to vowels of languages with smaller inventories.” (Jongman et al. 1989: 239f).

An interesting approach is presented by Carré (1996). He modelled vowels by a stepwise deformation of the acoustic tube (front to back constriction and labial command) and plotted the results in an F1/F2 plane, whereby F1 is on the abscissa and F2 on the ordinate. With this representation, he seemingly approached greater articulatory realism. The location of the constriction arises from a combination of F1

---

<sup>41</sup> Liljencrants & Lindblom (1972) and Lindblom (1986) offer various explanations for the unpredictability of vowel systems, especially that vowels systems are not determined by perceptual contrast alone. They do, however, not consider the possibility that perception of contrast is language-specific and, therefore, learned. Recent studies prove a connection between a speaker's perceptual acuity and his or her production of contrasts (see e.g. Perkell et al. 2004, 2006). Moreover, diachronic changes of vowel systems are not incorporated in the model either.

and F2; a maximally high F1 combined with a low F2 indicates a constriction location with the lowest  $X_c$  from the glottis. This would denote the pharyngeal constriction location for the vowel /a/. The other dimension captured by this representation is labialisation: a maximally low F1 and a maximally low F2 denote a labial back vowel whilst the labial front vowels exhibit a lower F1 and F2 than their respective unrounded cognates. In this approach, /ɛ/ and /œ/ can be represented as a back or as a front vowel. Similar results have been obtained by Boë et al. (1992) for /œ/. For this vowel, “the vocal tract most resembles a cylindrical tube” (1992: 35), and “the area functions reveal two approximately equal and symmetric minima around  $X_{c,g}^{42} = 5$  cm and  $X_{c,g} = 11$  cm” (1992: 36). As concerns the unrounded vowel, it is /æ/ rather than /ɛ/ which is described as having a narrowing just above the glottis (Fant 1980 = 2004, Wood 1979, Fant & Båvegård 1997). Fant & Båvegård (1997) illustrate that this vowel can be considered either as a front vowel with a wide constriction area or as an extreme back vowel. Two typical versions for this vowel are (from: Fant & Båvegård 1997: 7, Fig. 7):

/æ/	$X_c^{43}$	$A_c$	F1	F2	F3
Front configuration	4 cm	6 cm <sup>2</sup>	648	1595	2450
Back configuration	13.5 cm	2 cm <sup>2</sup>	654	1588	2452

In Figure 3.3, the data presented in Figure 3.1 are plotted in the same way as proposed in Carré (1996):

<sup>42</sup>  $X_{c,g}$  = constriction coordinate from the glottis.

<sup>43</sup>  $X_c$  = constriction coordinate from the incisors,  $A_c$  = constriction area

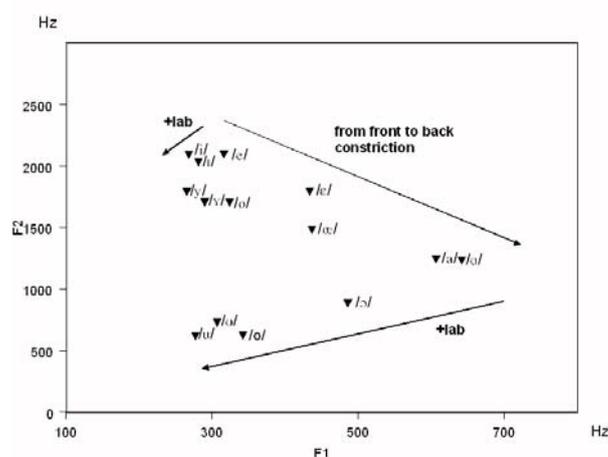


Figure 3.3: F1/F2 plot of the vowels of Standard Austrian German as proposed in Carré (1996), speaker sp012, logatome reading task.

In this model, the labial command always results in a lower F1 for the labial vowel. This does, in fact, not hold for the front labial vowels, as can be seen from Figure 3.3 and as has also been exemplified by Wood (1986). It seems that constriction location is not accurately modelled either, since F3 also substantially contributes to the determination of constriction location, especially in the /i/ and the /u/-vowels.

The non-linear relationship between acoustic data and articulation, as shown above for the vowel /æ/, was also known to Joos (as well as the non-conformity of the IPA vowel quadrilateral with the X-ray data) – “Two vowels might sound different and yet have the same two formants (...), but the reverse is not possible: if the formants differ, the sounds are not alike” (1948: 61) – however, despite his observations he seemed to have been fascinated by the possibility of correlating F1 and F2 with tongue height and backness respectively. Up to date, vowels are represented in a two-dimensional, mostly F1/F2 plane and interpreted in the way proposed by Joos (see e.g. Lindblom 1986, Disner 1986, McRobbie-Utasi & Starks 2001, Nowak 2006). For small vowel systems, or better, for vowel systems whose most peripheral constriction location on the front end of the vocal tract is indeed the hard palate, as stated in the IPA Handbook, the correlation of F2 and backness can be upheld as long as the degree of constriction is sufficiently small, preventing acoustic coupling of the front and the back cavity as well

as possible. This holds generally for the so-called “tense” vowels, i.e. /i/ and /e/, in vowel systems which lack front rounded vowels. As soon as the cavities are coupled through widening of constriction degree, F2 becomes more a function of constriction degree than of backness (Carré 2004, Beckman et al. 1995, Hoole 1999). In other words, a lowering of F2 in front vowels does not necessarily point to a centralization or retraction of the constriction location, nor does a rising F2 value of the back vowels necessarily point to a centralization or fronting of the constriction location, as the concept of Joos suggests and as has been assumed in many studies. If this concept of centralization were adopted for Standard Austrian German, /i/ of speaker sp129 in Figure 3.4 would have a more retracted tongue position with respect to /e/, an interpretation that is definitely wrong.

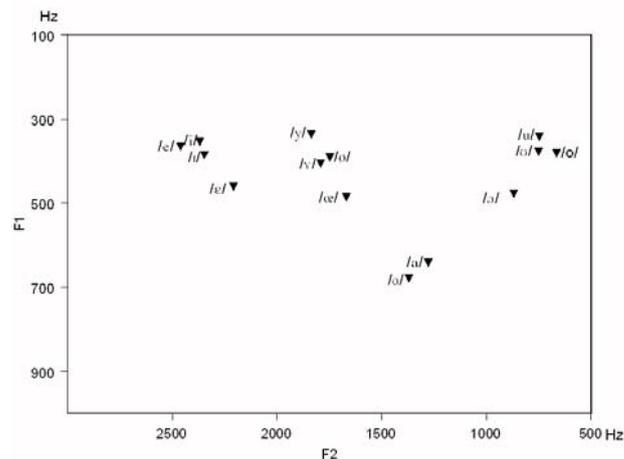


Figure 3.4: F1/F2 plot of the vowels of Standard Austrian German spoken by speaker sp129, sentence reading task.

A brief look at the mean values of F3 of /i/ and /e/ (Table 3.1) reveals that /i/ has, on the contrary, a more fronted constriction location than /e/. As soon as the constriction location transgresses the critical point of about 2/3 of the vocal tract length, a shift in cavity affiliation of F2 and F3 takes place, which causes F2 – now associated with the cavity behind the constriction – to drop or at least to stay constant and F3 – associated with the cavity in front of the constriction – to rise substantially (see e.g. Fant 1970,

2004, Stevens 1999, Johnson 1997, Badin et al. 1990, Gay et al. 1992). This shift in cavity affiliation is responsible for the equal or higher values of F2 for /e/ in Figures 3.1, 3.2 and 3.4 and might also be responsible for the same results in Iivonen (1987b).

	Fig. 3.1 Sp012		Fig. 3.2 Sp180		Fig 3.4 Sp129		Iiv. men		Iiv. women	
	/i/	/e/	/i/	/e/	/i/	/e/	/i/	/e/	/i/	/e/
<b>F1</b>	269	315	240	335	357	367	234	326	261	386
<b>F2</b>	2074	2092	2533	2566	2374	2458	2287	2256	2738	2712
<b>F3</b>	3263	2797	3410	3189	3113	3001	3103	2818	3423	3276

Table 3.1: Median<sup>44</sup> values for F1, F2 and F3 for the speakers represented in Figure 3.1 (sp012), 3.2 (sp180) and 3.4 (sp129) and mean values of F1, F2 and F3 for the data presented in Iivonen 1987b.

Assuming F3 to be a quarter wavelength resonance of the front cavity for /i/ and F2 to be a quarter wavelength resonance of the front cavity for /e/, the calculated constriction locations would be 2.7 and 4.2 cm from the incisors respectively for speaker sp012, and 2.6 and 3.4 cm from the incisors respectively for speaker sp180, distances which are in accordance with the measurements presented in Fant (2001) for Swedish vowels. Therefore, for the front vowels, Standard Austrian German discerns two constriction locations, and this result cannot be accounted for by a traditional F1/F2 representation.

As concerns the negative correlation of F1 with vowel height<sup>45</sup>, physiological realism is not met either, since e.g. the tongue has in most cases a higher position for /a/ than for /ɔ/ and even /o/ (Wood 1987, Bohn et al. 1992, Hoole & Kühnert 1996, Hoole & Mooshammer 2002).

Apart from the fact that an F1/F2 representation cannot capture articulatory adjustments in a satisfactory way, it additionally suggests that vowel articulation is gradual, i.e. a vowel can be articulated anywhere within the defined extreme points of the vowel quadrilateral. This is exactly the position the Handbook of the IPA takes:

<sup>44</sup> The median has been chosen in order to meet the high standard deviation of F3 for sp129, which is caused by the less formal task of reading sentences as compared to reading logatomes. For sp012 and sp180, there are hardly any differences between mean values and median, whereas for sp129, F3 shows substantial differences.

<sup>45</sup> Kohler (1998) uses “openness” as correlate with F1.

“Since the vowel space is continuous, it is a matter of chance whether a vowel in a language exactly coincides with one of the reference points symbolized on the quadrilateral. In particular, languages may use vowels which are similar to, but not as peripheral as, the reference points indicated by the cardinal vowels.” (Handbook IPA 1999: 13)

This statement leads to the question whether constriction locations of vowels are discrete or whether any location of the vocal tract is exploited for forming a constriction.

### 3.2. Constriction locations for vowels: discrete or gradual?

“The classical vowel model, originally introduced by Bell in 1867 (4) and modified into various versions by other authors, is characterized by the class of central vowels. The model was designed around the single resonance theory, according to which the upper surface of the tongue narrows the mouth channel locally in order to delimit the buccal cavity and tune its natural resonance. Bell postulated a configurative aperture that “may be shifted to any part of the back or front of the palatal arch” (p. 71). He held that the horizontal and vertical position of the tongue arch relative to the roof of the mouth set the size a location of this aperture, so that the natural resonance of the mouth cavity would rise progressively as the tongue moved from low to high at the back, central and front locations in turn.” (Wood 1987: 53)

This model acts on the assumption that tongue movements are gradual and constrictions can, therefore, take place anywhere on the roof of the mouth. Since a constriction can take place anywhere, central vowels are incorporated into this model. However, investigations based on x-ray studies suggest three or four discrete places of constriction<sup>46</sup>, according to the respective classification. Stratka (1978) classifies the vowels in three categories: alveolar vowels, which include the i-vowels, the y-vowels, the e-vowels, ø and schwa, pharyngeal vowels, which include the a-vowels, the o-vowels, œ, nasalized œ, ε, a, ɔ, and schwa<sup>47</sup> and velar vowels, which include the u-vowels.

The x-ray studies performed by Wood (1979, 1982) on several languages yielded 4 prominent constriction locations: a palatal one, a velar one, an upper pharyngeal one and a lower pharyngeal one. These tongue gestures are available for both vowels and

<sup>46</sup> The highest point of the tongue as reference point is abandoned.

<sup>47</sup> The schwa appears twice, because this vowel can either take a constriction in the pharyngeal or in the front region.

consonants (Wood 1996): as far as the central vowel is concerned, Wood states: “The articulation of so-called central vowels obviously needs clarification” (Wood 1991b: 64). Fant (2001), on modeling Swedish vowels, assumed 3 areas of constriction:

- a “front” region of  $X_c$ <sup>48</sup> located less than 4 cm from the teeth
- a “mid” region at coordinates between  $X_c = 4$  cm and  $X_c = 7$  cm and a
- a “back” region at  $X_c$  greater than 7 cm. (Fant 2001: 45f)

The region of midvowels was designed to provide a suitable transition between the very different front and back regions (Fant 2001), which is necessary for sequences like e.g. [ja], where a neutral state is involved but not completely reached (Lin & Fant 1989). For the Swedish vowels, with the exception of [u] at  $X_c = 6.5$  cm from the teeth (Fant 2001) and the neutral vowel [ə] at  $X_c = 6.4$  cm from the teeth (Fant & Båvegård 1997), no vowel was found well in the mid-range. The central vowels [ɨ] and [ɵ] have moved towards the front and the back respectively. The constriction location for [ɨ] is even more front than the one for [i] ( $X_c = 2.8$  cm vs.  $X_c = 3.1$  cm from the teeth respectively), and [ɵ] has its location near [o] ( $X_c = 8.8$  cm vs.  $X_c = 8.2$  cm respectively) (Fant 2001). On the whole, Fant’s results are in agreement with the constriction locations spotted by Wood (1979).

Russian is described as having a high central vowel [ɨ] as well. The area function<sup>49</sup> created in Fant (1980 = 2004) shows a constriction at about the same location as for [i] which is, however, wider and shorter. The difference between [i] and [ɨ] lies in the specific cavity affiliation of the first and second formant: in [i], F2 is affiliated with the cavity behind the constriction and F3 with the cavity in front of the constriction, whilst in [ɨ], the cavity affiliations are reversed, F2 is affiliated with the cavity in front of the constriction and F3 with the cavity behind the constriction.

---

<sup>48</sup>  $X_c$  is the constriction coordinate.

<sup>49</sup> It has to be considered that an area function is not in itself an articulatory parameter (Boë et al. 1992, Wood 1991a)

In the same way as in Fant (2001), schwa-vowels are described as either front or back vowels by Stratka:

“Pour la voyelle ə, la langue semble pouvoir se placer, selon les idiomes et les locuteurs, d’un côté ou de l’autre de la limite entre les voyelles alvéolaires et les voyelles pharyngales. En anglais le ə apparaît tantôt alvéolaire, tantôt pharyngal (...), tandis que dans d’autres langues, il est plutôt pharyngal: ainsi en allemand (...), en catalan (...), en bulgare (...), en chinois (...), et peut-être aussi en français (...).” (Stratka 1978: 450)

The so-called “indeterminate” vowel of Bulgarian, which is often denoted as /ə/ is described as exposing the same low pharyngeal tongue body gesture as /a/, but with a small mouth opening (Wood & Pettersson 1988, Wood 1996). The vowel is therefore characterized as “close pharyngeal”.

It seems that constriction locations in the mid range of the vocal tract are avoided and that at least phonological central vowels have an active constriction gesture either in the front or in the back of the vocal tract. Given the high and context-independent variability of both phonemic and allophonic “central vowels” (e.g. Eastern Arrernte, see Ladefoged & Maddieson 1996, or Albanian, see Granser & Moosmüller 2002, Moosmüller & Granser 2003, 2006), it has been proposed that central vowels are not specified for constriction location (Bates 1995) and can consequently take either a front or a back configuration. Schwartz et al.’s (1997) evaluation of the UPSID phoneme inventory concludes that phonemic “/ə/ is a “parallel” vowel which exists because of intrinsic principles (probably based on vowel reduction) different from those of other vowels” (p. 251) and does not seem to interact with other vowels. As a consequence, its presence or absence should not modify the structure of the vowel system. This seems to be the case in Québécois, where the schwa has merged with [œ] (Martin 1998).

The observation that allophonic schwa vowels exhibit a high level of context-dependency (van Bergem 1994, Bates 1995) and that they are more readily assimilated (e.g. in Danish, Jensen 2001) or deleted (Gheg variety of Albanian, Camaj 1969) than other unstressed vowels, led to the concept that schwa vowels are vowels without a target (van Bergem 1994) or that they are vowels with an active gesture that is

overlapped by the gesture of the following full vowel (Browman & Goldstein 1992). However, Hála (cited after Stratka 1978) could show that the schwa in English has either an alveolar or a pharyngeal constriction. More recent studies (Gick et al. 2000, Gick 2002) could prove a pharyngeal constriction for American English schwa. In sequences where adjacent phonemes exhibit conflicting articulatory targets (e.g. a tongue root advancement followed by a retraction of the tongue root), the tongue passes through a neutral space, rendering a schwa-vowel (Lin & Fant 1989, Gick & Wilson 2001, 2005). But the concept of targetless schwa cannot be completely abandoned: Davidson & Stone (2003) could demonstrate that in cases where phonotactically illegal consonant clusters have to be dissolved, the tongue does not coarticulate with the epenthetic/excrescent<sup>50</sup> schwa, but with the following consonant. However, an excrescent schwa is perceived. They conclude:

“the production of the preceding consonant...it has been shown that speakers’ tongue motion during their production of /zC/ sequences is not consistent with movement toward a **schwa** target.” (Davidson & Stone 2003)

These examples demonstrate that a vowel<sup>51</sup> can, in principle, be articulated without a target. However, where a constriction is intended for the production of either an epenthetical schwa, an allophonic schwa or a phonemic central vowel, it seems to be located either in the front part of the vocal tract, or in the pharyngeal region. Therefore, it might be useful to differentiate between a “neutral vowel (configuration)” and a “schwa-vowel”.

Since the central region of the vocal tract does not seem to be a preferred target for a constriction for vowels, it can be assumed that the tongue does not move gradually from front to central or from back to central, but that the tongue aims at discrete locations for the articulation of vowels. Consequently, changes in formant frequencies,

---

<sup>50</sup> Davidson & Stone (2003) differentiate between “epenthetical schwa”, i.e. a schwa having a target, and “excrescent schwa”, i.e. vowel without target.

<sup>51</sup> It might be perhaps confusing to name a truly excrescent vowel “schwa”.

especially F2 changes, do not point to a shift in constriction location, and the concept of “centralization” does not meet articulatory realism.

### 3.3. The acoustic-articulatory relationship

Paradoxically, it was also Joos (1948) who pointed at the non-linear relationship between acoustic output and articulatory adjustments (i.e. different articulatory adjustments can trigger the same acoustic output). In 1972, Stevens put forward his widely discussed “quantal theory of speech”. This theory not only gives evidence for the non-linearity of articulatory-acoustic mapping, but also links these observations to perception. Stevens (1972, 1989) found three zones – a palatal, a velar and a pharyngeal zone – , where vowel spectra are relatively insensitive to small displacements in constriction location, whereas in other regions of the vocal tract, small displacements render drastical spectral changes. These detected stable areas, which are in good congruity with the results of Wood (1979), ensure not only a relative stability of the formant frequencies with respect to displacements in constriction location, but constrictions in these areas additionally induce a narrow spacing of two spectral peaks, which – according to Stevens – lead to perceptual enhancement.

For example, the location of constriction for the low vowels is situated in the lower pharyngeal region. The low vowels strive either for a proximity of F1 and F2 with a constriction located about 7 – 9 cm from the glottis rendering the back vowel /a/, or for a proximity of F2 and F3 with a back cavity length of about 4 cm rendering the front vowels /a, æ/.

For the non-low front vowels, this perceptual enhancement is achieved through narrow spacing of F2 and F3, or F3 and F4. In varying the length of constriction (5 cm vs. 6 cm), Stevens in both cases observes

“a broad maximum of F<sub>2</sub> for configurations having a back-cavity length in the range 6.5 to 9 cm. In this region where F<sub>2</sub> is a maximum, this formant is relatively close to F<sub>3</sub>. When the constriction is even farther forward, F<sub>3</sub> becomes close to F<sub>4</sub>, while F<sub>2</sub> remains relatively high.

The exact location in the maximum in  $F_2$  and the distance between the formants in this cluster of  $F_2$ ,  $F_3$ , and  $F_4$  depend on the length and cross-sectional area of the constriction between the tongue dorsum and the hard palate. When the length of the back cavity decreases to the left of the  $F_2$  maximum in Fig. 8, there is a substantial decrease in  $F_2$ , and  $F_2$  becomes quite sensitive to changes in  $l_1$ <sup>52</sup>.” (Stevens 1989: 11)

The non-low back vowels are usually accompanied by lip-rounding. The reason for rounding is to bring  $F_1$  and  $F_2$  closer together as would be the case without rounding. Furthermore, rounding ensures a greater degree of freedom as concerns the placement of the constriction location:

“Another potential advantage of using a rounded configuration for non-low back vowels is that  $F_2$  passes through a minimum value as the position of the tongue body constriction is displaced along the upper pharyngeal and velar region of the vocal tract (...). When the tongue body is in this position that yields a minimum  $F_2$ , both  $F_1$  and  $F_2$  are relatively insensitive to changes in the constriction position. Thus the precision with which the constriction must be located to give a stable and low value of  $F_2$  is relatively lax.” (Stevens 1999: 290)

Contrary to Wood (1979), who discerns an upper pharyngeal and a velar constriction location for the /o/ and /u/ vowels respectively, Stevens (1989) unifies this area for acoustic reasons:

“The figure<sup>53</sup> shows a broad minimum for  $F_2$  over a range of length of the back cavity.  $F_2$  is within 100 Hz of its minimum value for  $l_1$  between 2 and 7.5 cm. Within this range of  $l_1$ , the spacing between  $F_1$  and  $F_2$  is 400-500 Hz, and, while  $F_1$  does not achieve a maximum value, it varies by only about 80 Hz.” (Stevens 1989: 13)

For the /u/-vowels two constriction locations are generally reported: a front configuration with  $X_c$  at about 6,5 cm (with a low  $F_2$  and a low  $F_3$ ) and a back configuration with  $X_c$  at about 10 cm (with a higher  $F_3$  and occasionally higher  $F_1$ ) (see Boë et al. 1992, Fant & Båvegård 1997, Wood 1979 on Southern British). The back configuration of the /u/-vowels has its constriction near /o/ ( $X_c = 11$  cm according to Fant & Båvegård 1997). This range of possible constriction locations for the /u/-vowels justifies a unification of Wood’s velar and upper pharyngeal regions. The fact, however, that this whole range is a quantal region, does not mean that the whole region is exploited for constrictions. For the /u/-vowels, two locations have been spotted so far. This leaves the

---

<sup>52</sup>  $l_1$  = back cavity length.

<sup>53</sup> i.e. the nomogram calculated for the configuration of a non-low back rounded vowel.  $l_1$  = back cavity length.

question of whether constrictions are intended in the region which separates the upper pharyngeal region for the /o/ and occasionally the /u/-vowels from the velar region for the /u/-vowels.

“Published nomograms give the magnitude of a formant shift that can be attributed to a gesture, although the three-parameter models [...] are difficult to understand in gestural terms, as they are really models of the area function, and not of the manoeuvres that created it.” (Wood 1991a: 215).

Steven’s quantal theory implies that articulation strives for a maximum of freedom in the positioning of the articulators. This implication is in contradiction with Wood (1982) who proved that all vowels are articulated with precision. Wood’s results as concerns articulatory precision have been confirmed by Hoole & Mooshammer (2002) in their work on German vowels and by Boë et al. (1992) for American English vowels. Therefore, the observation of acoustically stable regions need not necessarily imply that the whole of this stable region is utilized in production. Given the observed precision in vowel articulation, it is at least just as comprehensible that speakers utilize the possibility of varying the acoustic output by the means of small displacements of the articulators. The Distinctive Region Model (DRM; Mrayati et al. 1988, Carré & Mody 1997, Carré 2004) makes use of the principle of creating a maximum acoustic contrast between two sounds for a minimum area deformation (Carré 2004: 230). This, on first view, is the exact opposite of the quantal theory:

“...it has to be noted that a characteristic of stability in the articulatory-acoustic relation is the exact opposite of the characteristic of least effort, and thus of efficiency.” (Carré 1996: 434)

The Distinctive Region Model (DRM) is based on the sensitivity analysis of formant frequencies in vocal tract constriction (Fant & Pauli 1974). The distinctive regions are specified by the zero-crossings of the sensitivity function of a specific formant. Dependent on how many formants are incorporated into the model, the vocal tract is divided in either 2, 4 or 8 distinctive and symmetrical regions of unequal length. The constriction is located preferably in the middle of each region in order to produce the

most distinctive formant transitions. This also means that the vocal tract is quantal in its nature.

It has been put forward that the DRM model implicitly assumes that a linear relation between articulation and acoustic output exists (Boë & Perrier 1990). Linearity is, however, more a goal of the model than an assumption; the constriction locations are placed at the midpoints of each region, where transversal displacements guarantee maximal formant frequency changes, so as to optimize the principle to achieve maximal contrast with minimal area deformation (least effort principle):

“An important property of the model is that variation in the values of the regions’ cross-sectional areas around the uniform-tube configuration generates maximal formant frequency variations. In other words, if we look for places in the vocal tract having the best modulation of formant frequencies around the uniform configuration, and the largest dynamics of these modulations, these places are around the midpoints of the regions. Nonlinearities and saturations occur only near the borders of the vocalic space.” (Carré & Mrayati 1991: 436f)

Areas of stability, in this model, are not exploited in production. In order to examine whether areas of stability are exploited or not, one has to know which area functions result in an identical acoustic output. Atal et al. (1978) introduced the concept of fibers, which defines vocal tract shapes with identical acoustic properties. The authors conclude:

“Large changes in the shape of the vocal tract can be made without changing the formant frequencies. These changes are consistent with the hypothesis that compensatory articulation is a possibility—that is, different people can produce the same sound with different vocal-tract shapes.” (Atal et al 1978: 1555)

Although “an area function is not in itself an articulatory parameter” (Boë et al 1992: 29, but see also Wood 1991a), it is possible to derive to a certain extent vocal tract adjustments from the speech signal (Boë et al 1992, Ladefoged et al. 1978). Boë et al. (1992) reduced the possible area functions to meaningful configurations and could classify 7 out of 10 vowels uniquely. The vowels [i, e, a, y, o, ɔ] show a relatively precise control of the constriction location and of constriction degree. [i, y, e] show a strong constraint and maintain a small constriction degree ( $A_c < 1 \text{ cm}^2$ ). The vowels [ɛ,

ø, œ] exhibit a large variation in constriction degree (from 0.2 cm<sup>2</sup> to 4.0 cm<sup>2</sup>). The authors conclude:

“The parameters  $X_c$ ,  $A_c$  and  $A_1$  thus seem to be good candidates for use in inversion procedures. It must be noted though that the precision of control necessary for each parameter depends on acoustic sensitivities, which vary according to the vowel under consideration (...). Thus [i] requires precise control of  $X_c$ , but no precision at all for  $A_1$ . The case of [u] is different, requiring a very small and accurate lip opening, but granting more latitude for  $A_c$ .” (Boë et al. 1992: 36)

Again, the degree of freedom guaranteed for either  $A_1$  in [i] or  $A_c$  in [u] need not be reflected in actual articulatory adjustments. On the contrary, Wood (1982: 46) could show that  $A_1$  is precisely controlled in the front vowels: the jaw opening is smaller than 8,9 mm for the i-vowels, but greater than 8,9 mm for the e-vowels. These results were confirmed by Hoole & Mooshammer (2002).

Regions of stability, which in principle allow for a certain degree of freedom of articulatory adjustments, exist. Given the high degree of precision in the articulation of (vowel) phonemes, it is not quite clear what their function is. In quantal theory, regions of acoustic stability are separated by regions of acoustic instability, and these natural boundaries define the opposition between distinctive features (Stevens 2003). The boundary, e.g. between [+back] and [-back] vowels, occurs at a location where F2 is near the lowest frequency of the subglottal system and the interaction between the sub- and supraglottal resonances cause discontinuous jumps in the spectrum (Stevens 2003). This is certainly a conclusive explanation for the rough division of [+back] and [-back] vowels. Such a natural boundary is not always provided by quantal theory however (e.g. within the front region). On the contrary, quantal theory predicts a high stability of the vowel /i/. This has already been confuted by Abry et al. (1989). The current investigation reveals that in Standard Austrian German, the /i/ vowels are located exactly at or slightly before such a natural boundary where formant frequencies converge and a slight displacement of the tongue position or a reduction in constriction length causes a considerable change in formant frequency values, especially F3. Therefore, the stability

of the /i/-vowels only holds for languages which do not expose a high amount of front vowels, including front rounded vowels. like e.g. English.

Acoustic stability is, therefore, not necessary for the creation of vowel systems (see also Lindblom 2003), although a positive correlation between the locations of vowels from a small vowel inventory and the regions of stability might be observed. Yet, the vowels of smaller vowel inventories are articulated with no less precision than the vowels of greater inventories (Jongman et al. 1989, Flege 1989). I.e. the potential freedom of movement is not exploited in small inventories either. As soon as the number of vowels exceeds the number of the most frequent vowel system in the languages investigated so far (Maddieson 1984, Ladefoged and Maddieson 1996), namely five, further locations or adjustments have to be exploited in order to gain distinctiveness. Jaw opening, for example, can be combined with constriction degree in such a way that four distinct vowels can be produced (Wood 1982):

	/i/	/ɪ/	/e/	/ɛ/
open	–	–	+	+
constricted	+	–	+	–

A displacement of the constriction location is not necessary for the creation of these four phonemes. Standard Austrian German, however, discerns eight phonemes in the front region. As has already been pointed out by Wood (1982, 1986), languages with front rounded vowels exploit an additional pre-palatal region to differentiate the /i and y/-vowels from the /e and ø/-vowels, and this pre-palatal region is highly sensitive to small articulatory displacements. Therefore, the more vowels that have to be discerned, the more problems they pose for vowel models.

Regions of acoustic stability might be useful in allophonic variation, in order to guarantee perceptual invariance. However, Stevens' regions of stability only refer to constriction locations and not to degree of constriction. Allophonic variation affects the degree of constriction rather than the constriction location, and variation in constriction degree renders a monotonic change in formant frequency values (Gay et al. 1991, 1992,

Carré 2004). The overshoot and undershoot simulations performed by Gay et al. (1991, 1992) clearly show that shifts in perception for /i/ were more dependent on  $A_c$  than on  $X_c$  and shifts in perception for /u/ were highly dependent on  $A_l$ . Therefore, it can be concluded that quantal regions neither serve the creation of phonological oppositions nor are exploited in allophonic variation. Consequently, one can ask, what are they good for?

Inter-speaker differences in articulation have been documented in many investigations. It is assumed that these inter-speaker differences render different acoustic output. E.g. Apostol et al. (1999) report inter-speaker differences which are caused by speaker specific strategies used to control the speech apparatus. In the production of the vowel /a/, one of their speakers exposed a short back cavity and a long constriction zone, whereas the other speaker had a short constriction, but a rather long back cavity. These articulatory differences are of course reflected in the acoustic output, which excelled by a higher F2 for the first speaker. In a similar way one speaker had a long front cavity in the production of the vowel /i/, whose quarter-wavelength resonance is F2, whereas the other speaker had a very short front cavity, but a long constriction zone (Apostol et al. 1999: 446).

Maeda (1991), however, reports that articulatory variability was higher than acoustic variability in the production of French /a/ for *pâte* (pastry) vs. /a/ for *à* (PREP., to) (two vowels which have merged acoustically in some modern French variants). Whereas formant frequencies measured at the center of the vowels were quite similar, articulatory differences could be observed: the jaw was clearly open during the vowel /a/, whereas during the vowel /a/ the jaw was at an average position. The vowel /a/ exhibited a falling pattern (i.e. a fronting gesture) of the tongue dorsum position, the vowel /a/ indicated a peak (i.e. backing followed by fronting). These results indicate not only that a considerable articulatory variability for the same vowel exists, but also, and

more importantly, that articulatory gestures are applied in a way that compensates articulatory differences in order to produce invariant acoustic patterns.

The most obvious differences in vocal tract shape is that between children, women and men. These differences should lead to different formant frequency values. However, despite the difference in vocal tract length, F1 and F2 of the vowel /u/ and F3 of the vowel /i/ of women are very similar to those of men (Fant 1980=2004). Fant states that differences in perceptually important formants are minimized by compensations.

“Differences in perceptually important formants may thus be minimized by compensations in terms of place of articulation and in the extent of the area function narrowing. Such compensations are not possible for all formants and cannot be achieved in more open articulations. The great difference in F2 of [i] is in part conditioned by the relatively short female pharynx but can in part be ascribed to the retracted place of articulation.” (Fant 2004: 44).

The differences in vocal tract shape between men and women are of course an extreme example, but even here compensatory strategies can be observed. Between the sexes, slight differences in vocal tract shapes have to be assumed, and these slight differences may be compensated. Regions of stability ensure or facilitate such compensations. From this perspective, it makes sense that constriction locations are less sensitive to small displacements, whereas displacements in constriction degree yield monotonous changes in formant frequencies.

## 4. Vowel inventory and features

Descriptions of the vowel inventory (without diphthongs) of Standard German set the number of vowels at between 17 (Wiese 1996), 16 (Jørgensen 1969, Wurzel 1970, UPSID, see Simpson 1998, 1999, Kohler 1999), 15 (Moulton 1962, Sendlmeier 1985, Iivonen 1987a, b), 14 (Heike 1961<sup>54</sup>, Rausch 1972) and 8 vowels (Becker 1998). The reason for this inconsistency in range lies in the different views of the status of the schwa and long /ɛ:/. Jørgensen (1969), Kohler (1999) and UPSID assume a mid-central schwa. Wiese (1996) additionally assumes a phonemic status for /ɐ/, the vowel resulting from r-vocalization. The descriptions with more than 15 vowels all include the long /ɛ:/<sup>55</sup> in their analysis. Moulton (1962) rejects the long /ɛ:/, but includes a mid-central schwa. Iivonen (1987a) includes the long /ɛ:/, but assumes no mid-central schwa. In his analysis of Standard German, /ɛ:/ nearly merges with /ɛ/. In Northern German, however, /ɛ:/ merges with /e/ (Heike 1961, Jørgensen 1969, Kloeke 1982). All analyses of Standard German assume two a-vowels, a long and a short one. Becker (1998) assumes only 8 vowels; these can, however, appear as long or as short vowels, according to their position in the syllabic structure. The open vowels, according to his analysis, differ with respect to duration, the closed and mid vowels with respect to centralization; consequently, the short /e/ and /ä/<sup>56</sup> merge.

---

<sup>54</sup> Heike (1961) regards the division of sounds into “vowels” and “consonants” as inadequate, and prefers to distinguish “Kernphoneme” (nucleus phonemes) and “Satellitenphoneme” (satellite phonemes).

<sup>55</sup> see Becker (1998) for a thorough overview of the analytical status and historical development of long /ɛ:/.

<sup>56</sup> Becker (1998) assumes a different quality for short and long /ɛ/ and therefore adopts the symbol /ä(:)/ for the long vowel (1998: 23). Nevertheless, this qualitatively different vowel merges with the short /e/ (1998: 163).

For Standard Austrian German, the phonemic position of the schwa need not be discussed, since a mid central schwa exists neither phonetically nor phonologically<sup>57</sup>. As concerns /ɛ:/, Iivonen (1987b) still assumes this vowel to be a part of Standard Austrian German, although it strongly approaches /e:/. In the current investigation, /ɛ:/ has totally merged with /e:/, therefore it can be discarded as a phoneme. As a starting point, the phoneme inventory of Standard Austrian German is assumed to consist of the following 14 (underlined> vowels:

<b>bieten</b>	b $\underline{i}$ tɛn	<b>Hüte</b>	h $\underline{y}$ tɛ	<b>spuken</b>	ʃp $\underline{u}$ kɛn
<b>bitten</b>	b $\underline{i}$ tɛn	<b>Hütte</b>	h $\underline{y}$ tɛ	<b>spucken</b>	ʃp $\underline{u}$ kɛn
<b>beten</b>	bɛ:tɛn	<b>Höhle</b>	hø:lɛ	<b>Mode</b>	m $\underline{o}$ dɛ
<b>betten</b>	bɛ:tɛn	<b>Hölle</b>	hø:lɛ	<b>Motte</b>	m $\underline{o}$ tɛ
<b>Schall</b>	ʃ $\underline{a}$ l			<b>Schal</b>	ʃ $\underline{a}$ l

Phonemically, these vowels are very unevenly distributed. A counting of all phonemes in spontaneous speech over all speakers resulted in a predominance of front unrounded vowels. The front rounded vowels hardly play any role, and the back rounded vowels occur relatively rarely as well (see Figure 4.1):

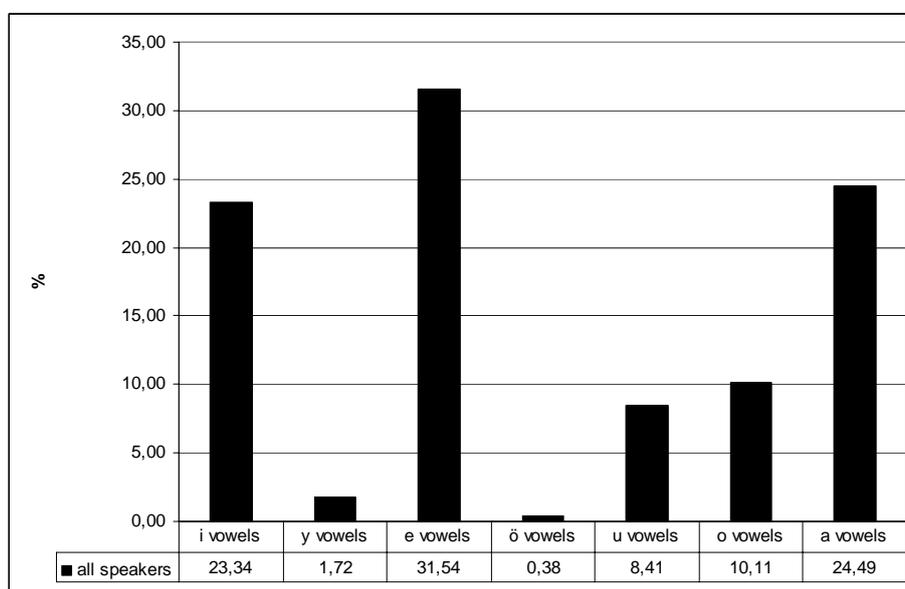


Figure 4.1: Occurrence (in %) of the vowel pairs in spontaneous speech, pooled over all speakers and all prosodic positions.

<sup>57</sup> For example, in pre-tonic /e/ of the prefix “ge-“, none of the speakers produces formant frequency values that would justify the assumption of a schwa.

Figure 4.1 reveals that the /e/ – vowels dominate the Austrian German vowel system, followed by the /i/ – vowels and the /a/ – vowels. Dividing the vowels according to two stress levels, the /e/ – vowels, with 21.9%, lose their predominant position for the stressed vowels and are relegated to the third place after the /a/ – vowels and the /i/ – vowels (with 26.3% and 25.5% respectively). In unstressed positions, however, they come first and, with 34.4%, outclass the /a/ – vowels and the /i/ – vowels by more than 10% (24% and 22.7% respectively). The /u/ – vowels and the /o/ – vowels are evenly distributed over stress positions: 9.8 % and 8% for the /u/ – vowels in stressed and unstressed positions, and 11.7% and 9.6% for the /o/ – vowels.

The 14 vowels are usually discriminated according to the following distinctive features:

Tongue height [±high]:	/i:, ɪ, y:, ʏ, u:, ʊ, o:, ɔ/ vs. /e:, ɛ, ø:, œ/ vs. /a, ɑ:/
Tongue position [±front]:	/i:, ɪ, y:, ʏ, e:, ɛ, ø:, œ, a/ vs. /u:, ʊ, o:, ɔ, ɑ:/
Lip position [±round]:	/u:, ʊ, o:, ɔ, y:, ʏ, ø:, œ / vs. /i:, ɪ, e:, ɛ, a, ɑ:/
Tenseness [±tense]:	/i:, y:, e:, ø:, u:, ʊ, o:, ɑ:/ vs. /ɪ, ʏ, ɛ, œ, ʊ, ɔ, a/

The last of the four categories has been and still is the subject of extensive controversy. First of all, it is not quite clear what the exact articulatory correlates to tenseness should be. Is it tenseness of the vocal tract walls, or tenseness of lip articulation triggering a more spread or more protruded articulation of the tense vowels as compared to the lax vowels? Is it tenseness of the tongue, resulting in a tighter constriction and a longer constriction area for the tense vowels as compared to their lax counterparts? Or does tenseness correlate with settings of the pharyngeal region, resulting in an advancement of the tongue root as has been described for several African languages like Igbo (Ladefoged 1964), Akan (Tiede 1996), Degema (Fulop et al. 1998), Maa (Guion et al. 2004)?

The observed opposition might also be the result of distinctive jaw position, the tense vowels being more closed than their lax counterparts. In this case, the feature

[±tense] would have to be substituted for the feature [±closed], as proposed by Stevens (1999).

A further opposition is evident from the distribution of the so-called tense vs. the so-called lax vowels: tense vowels are described as long, and lax vowels as short. The relevant distinctive feature could, therefore, also be a temporal one, namely [±long] (Hertrich & Ackermann 1997).

A brief look at the orthographic representation of items with tense or lax vowels shows that items with tense vowels are represented as a sequence of vowel + consonant, whereas items with lax vowels are represented as vowel + double consonant. This graphemic representation has historical roots and reflects the fact that Classical Middle High German made use of ambisyllabic consonants.

#### 4.1. The feature [±tense]: brief historical outline

A comprehensive historical outline of the opposition, its change and results, both in Standard German and two German dialects (Alemannic and Bavarian), has been given by Ronneberger-Sibold (1999). It will be summed up here briefly, since the results of her discussion shed light on many of the manifold observations made in vowel production.

Classical Middle High German made use of a three-way opposition as concerns the stressed syllables:

schāle	schale	Schalle
/ʃa:lə/	/ʃa.lə/	/ʃal.lə/
shell, dish	inflected form of <i>schal</i> , stale	to resound (inflected form)

With respect to prosodic length, two short syllables were equivalent to one long syllable, therefore, *schale* /ʃa.lə/ was equivalent in length with /ʃa:/ in *schāle* and with /ʃal./ in *schalle* (each consisting of two moras). These durational equivalences were an

important element of Old High German and Middle High German quantifying poetry; therefore, it was concluded that the distinctive feature was a temporal one.

This Middle High German three-way opposition was reduced to two-way oppositions, but in different ways for different variants:

In some Alemanic variants, the geminates were given up: *schalle* merged with *schale*. Two types of open syllables were left: /ʃaː-/ in *Schale* (*shell, dish*) with a long vowel and /ʃal-/ in *schalle* (to resound, inflected form) and *schale* (stale, inflected form) with a short vowel.

The reduction of oppositions was not as simple in the non-Alemanic variants. Firstly, in Standard German and in the Central and North Bavarian variants, the Middle High German items of the type /ʃa.lə/ with a short open syllable were given up. Items of this type merged either with the type /ʃaː.lə/ – a long open syllable – or with the type /ʃal.lə/ – a closed syllable. Therefore, as oppositions two types remained: /ʃaː.lə/ and /ʃal.lə/. The result was syllabic isochrony: all stressed syllables were two moras long.

This syllabic isochrony has been retained in Central and North Bavarian with an additional constraint concerning obstruents: a long vowel is followed by a lenis consonant, e.g. *Feder* (feather) [ʏeːdɐ] and a short vowel is followed by long or fortis consonant, e.g. *Vetter* (cousin) [ʏetɐ], [ʏeɖɖɐ] or [ʏettɐ], consonant and vowel forming a structural unit (see also Bannert 1977 for a thorough discussion of isochrony in Bavarian dialects). In this context, it is of interest that Swedish, Norwegian and Icelandic show the same sort of 'complementary quantity' (Schaeffler et al. 2002).

In Standard German, where all geminates were degeminated, the short vowels remained. The former closed syllables became open and the contrast between short and long vowels has been enhanced by a co-occurring contrast of quality. The Bavarian constraint of a vowel – consonant combination is, therefore, not found in Standard German. In this variant, all four possible combinations of vowels and consonants are found:

Long vowel + lenis plosive	Mieder (bodice)	/mi:der/
Long vowel + fortis plosive	Mieter (tenant)	/mi:tɛr/
Short vowel + fortis plosive	Mitte (middle)	/mitɛ/
Short vowel + lenis plosive	Widder (ram)	/vɪdɛr/

In Standard German, vowel + consonant do not form a structural unit. Ronneberger-Sibold concludes, supported by her analysis of the pronunciation of lexical creations, that the assumption of ambisyllabic consonants, and consequently the assumption of a closed syllable in items like *schalle*, has no evidence in Standard German. Standard German and the Alemanic dialects have to be analysed with what Ronneberger-Sibold describes as the plain model, i.e. the stressed syllables both in *Schale* (shell, dish) and *schalle* (to resound, inflected form) are open, the former containing a long vowel, the latter a short vowel. The ambisyllabic model, which requires syllabic isochrony, makes sense in an analysis of Central and North Bavarian dialects, where, at least in many variants, ambisyllabic consonants can be observed (for a more detailed discussion see Ronneberger-Sibold 1999).

## 4.2. Temporal Analysis

### 4.2.1. Isochrony in Standard German

As can be seen from the above analyses, the morae-counting structure of Classical Middle High German has developed in different ways in diverse variants of German, resulting in two main groups: those preserving isochrony and those giving up isochrony.

Standard German, as has been discussed in Ronneberger-Sibold (1999), belongs to the group that has given up isochrony. Instead, stressed syllables have either a long or a short vowel, without any structural constraints as concerns the following consonant. Moreover, Standard German has changed the quality of the short vowels, in this way maintaining, or better, enhancing distinctiveness. Jessen et al. (1995) showed that the distinction of vowel quality (based on F1 and F2 measurements) is maintained in both

stressed and unstressed conditions for the non-low vowels; vowel duration, however, is statistically significant only in stressed vowel conditions, but not in unstressed conditions. As concerns vowel duration, the low vowels behave in the same way as the non-low vowels (distinctive in stressed positions, not distinctive in unstressed positions), but show no difference in vowel quality. Jessen et al. (1995) conclude that the low vowels contrast in quantity, whereas the others contrast in quality. Their findings are in agreement with earlier studies (Ramers 1988, Iivonen 1984, Kohler 1995<sup>2</sup>). A later study based on the Kiel Corpus also corroborate these results (Kohler 1998). The perceptual experiments conducted by Sendlmeier (1981) proved the relevance of temporal information for the vowel pairs /a, a:/, /ɪ, e:/, /ʏ, ø:/ and /ʊ, o:/. In order to discriminate these pairs, temporal information was used by listeners in Sendlmeier's study, whereas for the discrimination of the remaining pairs, spectral information was applied. Nevertheless, on the basis of these experiments, Sendlmeier (1981, 1985) concludes that a primary quantitative opposition can without doubt be assumed for the pair /a, a:/, whereas for the vowel pairs /ɪ, e:/ and /ʊ, o:/,

“scheint es ebenfalls gerechtfertigt, eine phonologisch relevante Quantitätsopposition zu etablieren. Die die Differenzierung unterstützende Funktion eines vorhandenen Qualitätsunterschiedes spielt für diese Vokalpaare jedoch eine wichtigere Rolle als bei den beiden A-Lauten.” (Sendlmeier 1985: 194)

For the remaining pairs, spectral information was used for discrimination. Sendlmeier's results corroborate the hypothesis of Bennett (1968) on German and English, that the importance of duration is inversely proportional to the distance in quality a certain vowel pair exhibits: i.e. the smaller the distance in quality, the higher the importance of duration. An additional test on a pair of back, unrounded vowels [ʏ] and [ʊ:], which occur neither in German nor in English, revealed that German subjects used primarily spectral information for discrimination. Sawusch (1996), in his study on the perception of the vowels [ɛ] vs. [æ:] in “head” and “had”, could show that duration becomes an identifying cue when other sources of information were made ambiguous. The results of Strange & Bohn (1998) on the perception of North German vowels can be interpreted in

the same way: listeners discriminated tense and lax vowel pairs much better when temporal information was additionally available in the silence center stimuli<sup>58</sup>. The vowel chart presented in Sendlmeier (1981: 302) only displays spectral differences in one dimension (either F1 or F2) for those vowels, which were judged primarily for duration, whereas the other vowels, which were judged on the basis of spectral differences, differ for both F1 and F2.

In Swedish, a language which also discerns long/tense vs. short/lax vowels (Schaeffler & Wretling 2002), on the other hand, listeners contrast the vowel pairs /i:/-/i/ and /o:/-/ɔ/ primarily on the basis of durational information, whereas they use additional spectral information for the pair /ɑ:/-/a/ (Behne et al. 1999). Schaeffler & Wretling (2002), who compared long and short /a/ in Northern Swedish dialects, proved the hypothesis that stronger quality differences only go along with minor durational differences and vice versa for some dialects.

No spectral information is used for the discrimination of long and short vowel pairs in Japanese (Behne et al. 1999), although vowel duration is not robust across speech styles (Kozasa 2002) and spectral differences can be observed (Hirata & Tsukada 2004). In discriminating long vs. short vowels in Japanese, listeners make use of word/vowel ratios, which proved to be very stable across speech styles (Hirata 2004).

Therefore, the primary feature contrasting vowel pairs in a given language can either be a temporal or a spectral one. The respective other – secondary – feature might or might not be made relevant, when the other either fails to convey the required opposition or starts to become neutralized. Perception tests reveal that Japanese and Swedish contrast their vowels along a temporal dimension. The primary contrastive feature therefore is [ $\pm$  long]. In English and German, on the other hand, listeners primarily use spectral information to contrast the vowels. Vowel duration only becomes relevant when spectral information is blurred. Therefore, it can be concluded that

---

<sup>58</sup> The main aim of the Strange & Bohn (1998) study was to test the relevance of dynamic information in the perception of coarticulated vowels.

duration is only an accompanying, secondary feature. The same conclusion is drawn by Strange & Bohn (1998) for the high and mid vowels. The low vowels are treated differently. However, the results presented in Sendlmeier (1981) reveal that the relevance of temporal information does not depend on vowel height, but rather on the amount of spectral information available. This strongly points towards neutralization of certain vowel pairs, especially the low vowels. It can be assumed, in line with the historical analysis presented in Ronneberger-Sibold (1999), that durational contrasts are remnants of a former temporal organization. Therefore, it is argued that German contrasts its vowels with respect to quality, resulting in a vowel set of 13 vowels. Instead of assuming a different feature for the low vowels, /ɑ:/ and /a/ have been neutralized in this analysis. The results presented on the Standard Austrian German vowels will provide additional evidence for such an analysis.

#### **4.2.2. Isochrony in Central Bavarian**

According to Ronneberger-Sibold (1999), the Central and North Bavarian dialects maintain isochrony with the segmental constraint that a long vowel has to be followed by a lenis plosive and a short vowel has to be followed by a fortis plosive. The isochrony of the structural units “vowel+consonant” has been tested by Bannert (1977) and Ronneberger-Sibold (1999).

It is not the objective of the current investigation to test isochrony in the diverse variants of Central Bavarian. However, since the interaction between Standard Austrian German and Central Bavarian dialects (in particular the Viennese dialect) is quite strong, vowel and consonant durations of one Bavarian Dialect speaker (Upper Austria, region of Wels) have been analysed. This speaker was asked to read the same list of sentences as the Standard Austrian German subjects. A significant negative correlation ( $r = -0.5$ ,  $p = 0.00$ ) was observed. However, comparing the vowel and plosive durations, isochrony was observed for the group’s long vowel + lenis plosive and short vowel +

fortis plosive (see Figure 4.2), but not for the group's long vowel + fortis plosive and short vowel + lenis plosive. No statistically significant differences in duration were observed for the group's long vowel + lenis plosive and short vowel + fortis plosive. These results corroborate the hypotheses of Bannert (1977) and Ronneberger-Sibold (1999). The other pair lacks isochrony, because long vowel + fortis plosive and short vowel + lenis plosive are not possible combinations in Central Bavarian.

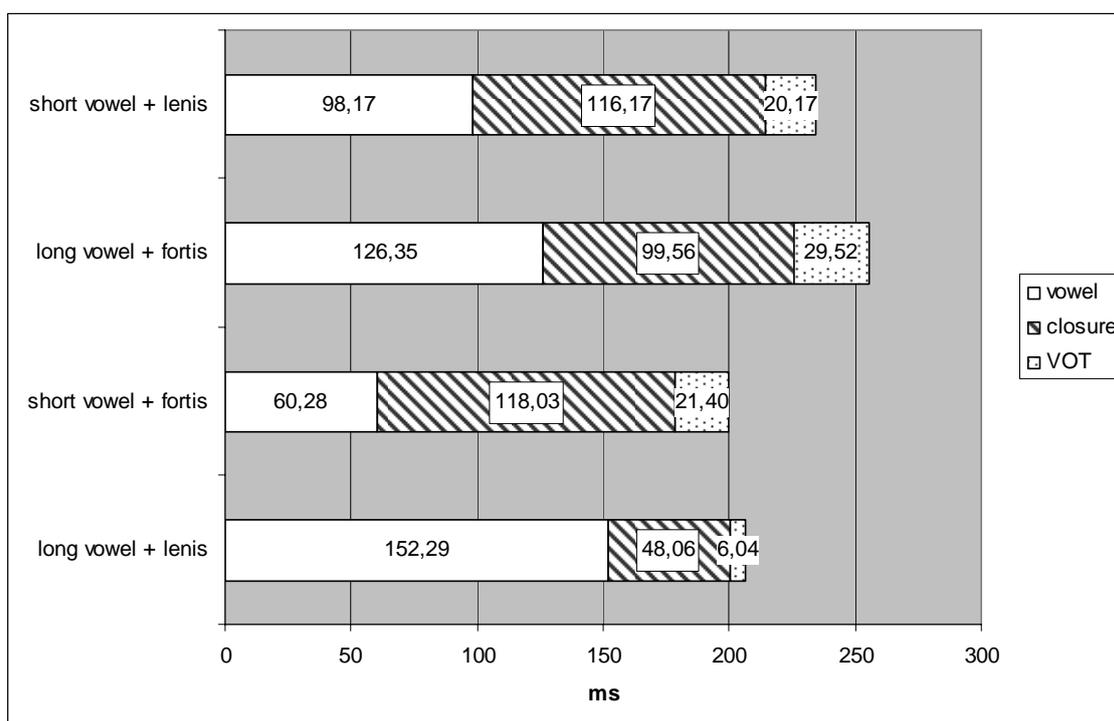


Figure 4.2: Duration of vowel, closure of the plosive and VOT of 54 bisyllabics, p182, sentence reading task.

In spontaneous speech, though, no isochrony can be observed. Long vowel + lenis plosive and short vowel + fortis plosive durations differ significantly and, most interestingly, the combination short vowel + fortis is longer than the combination long vowel + lenis plosive (see Table 4.1).

Spontaneous speech	Long vowel + lenis plosive	Short vowel + fortis plosive	n	t	p
Total duration (ms)	123.7	187.1	33	4.62	0.00
Plosive duration (ms)	36.0	111.5	33	12.89	0.00

Table 4.1: Mean durations of the total duration of vowel + plosive and of the duration of the plosives alone, and the statistical results. Speaker182, Central Bavarian, spontaneous speech.

Table 4.1 reveals that the duration of the plosive is responsible for the difference between the two vowel + plosive combinations. Long and short vowels do not differ to a statistically significant degree. Consequently, absolutely no correlation can be found between vowel duration and plosive duration ( $r = 0.16$ ,  $p = 0.32$ ): see Figure 4.3.

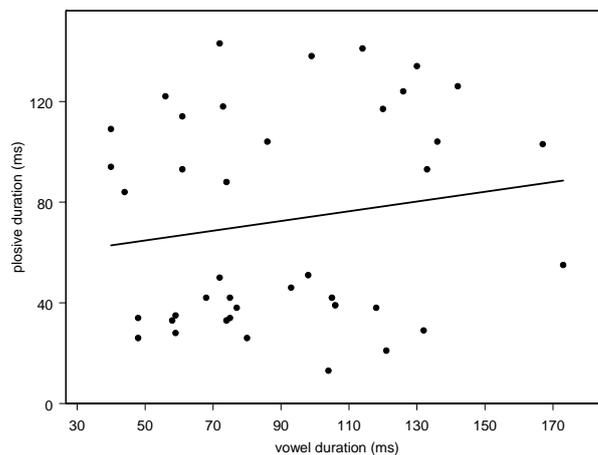


Figure 4.3: Plosive duration against consonant duration, speaker sp182, Central Bavarian, spontaneous speech.

Isochrony, for this speaker, is restricted to the very formal task of reading sentences. In spontaneous speech, durational aspects are reserved for other – overall timing – functions, and no longer serve to discriminate vowels or vowel + consonant combinations.

It has to be emphasized that these few measurements are not meant to reflect the situation of temporal organization of Central Bavarian vowels and consonants. Rather, they are meant to indicate that the existence of isochrony is not as secure as assumed,

and that further durational measurements on larger corpora, and especially on spontaneous speech, have to be performed in order to answer the question of whether quantity plays a distinctive role in Central Bavarian.

### 4.2.3. Isochrony in Standard Austrian German

Standard Austrian German has a difficult position. On the one hand, it models itself on Standard German, where all possible vowel + consonant combinations are conceivable. On the other hand, in phonology and prosody, it has Bavarian roots and is strongly connected with Central Bavarian.

In order to test whether the distinctive feature usually labeled as [ $\pm$  tense] in Standard German has to be analysed as outlined above in Standard Austrian German, the durations of vowel, closure of the following plosive, and VOT of bi-syllabics in stressed positions have been measured. All four possible combinations of vowel + plosive have been investigated:

- Long vowel + lenis plosive: type: *Liebe* (love), *Grabe* (grave: old DAT) – bilabial, *Lieder* (songs), *Made* (maggot) – alveolar, *Tiger* (tiger), *Lage* (circumstance) – velar.
- Short vowel + fortis plosive: type: *Lippe* (lip), *doppelt* (double) – bilabial, *bitte* (please), *Ratte* (rat) – alveolar, *Hecke* (hedge), *Hacke* (hoe) – velar.
- Long vowel + fortis plosive: type: *Siepe* (family name), *Hupe* (horn) – bilabial, *bietet* (offer, inflectional form), *rate* (guess, inflectional form) – alveolar, *Ekel* (disgust), *Luke* (hatch) – velar.
- Short vowel + lenis plosive: type: *Krabbe* (crab) – bilabial, *Widder* (ram), – alveolar, *Egge* (harrow) – velar.

110 tokens per person of the types described above of the sentence reading task of one female speaker (sp180) and one male speaker (sp127) have been analysed. Performed t-tests proved no statistically significant differences with respect to vowel duration (in

ms) and VOT between sp127 and sp180. However, statistically significant differences with respect to closure duration were observed. Therefore, the results have not been pooled together.

The results obtained for the Standard Austrian German speakers resemble those obtained for the Central Bavarian speaker: there is a significant negative correlation between vowel duration and plosive duration ( $r = -0.41$ ,  $p = 0.00$  for speaker sp180 and  $r = -0.58$ ,  $p = 0.00$  for speaker sp127), i.e. vowel duration does not increase with decreasing plosive duration. However, in the same way as for the Central Bavarian speaker, isochrony can be observed for the categories “long vowel + lenis plosive” and “short vowel + fortis plosive”. The remaining two categories do not, however, fit into this pattern (see Figure 4.4 and 4.5). Conducted t-tests proved no statistically significant differences between the durations of the categories “long vowel + lenis plosive” and “short vowel + fortis plosive” for both speaker sp180 ( $n = 70$ ,  $t = 1,36$ ,  $p = 0.09$ ) and speaker sp127 ( $n = 70$ ,  $t = 0.003$ ,  $p = 0.50$ )<sup>59</sup>.

In the same way as for the Central Bavarian speaker, isochrony is no longer present in spontaneous speech. Durations of the categories “long vowel + lenis plosive” and “short vowel + fortis plosive” differ considerably, and the category “short vowel + fortis plosive” is again longer than the category “long vowel + lenis plosive”, as becomes apparent from Tables 4.2 and 4.3.

---

<sup>59</sup> The category “short vowel + lenis plosive” shows no statistically significant differences either. Therefore, it is mainly the category “long vowel + fortis plosive” which does not fit. However, the former category plays only a marginal role in the lexicon of German, whereas the latter category is, from a quantitative point of view, highly important.

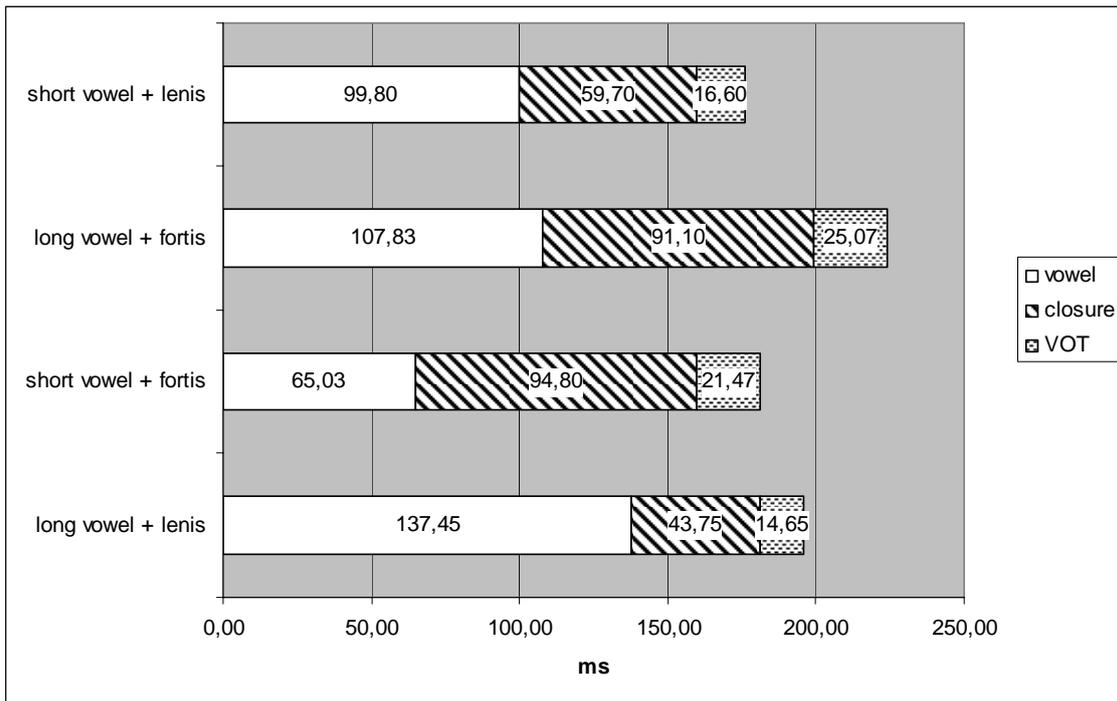


Figure 4.4: Durations (mean values in ms) of vowel, closure and VOT of 110 bi-syllabics, speaker sp180, sentence reading task.

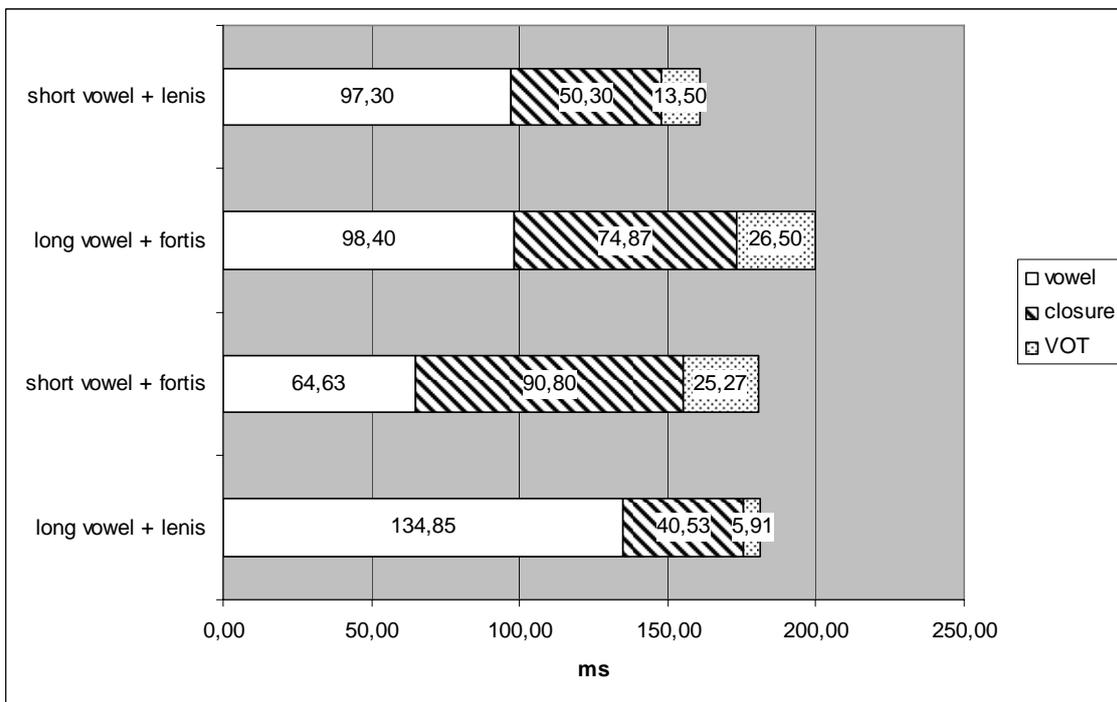


Figure 4.5: Durations (mean values in ms) of vowel, closure and VOT of 110 bi-syllabics, speaker sp127, sentence reading task.

Spontaneous speech	Long vowel + lenis plosive	Short vowel + fortis plosive	n	t	p
Total duration (ms)	139.9	201	21	3.25	0.00
Plosive duration <sup>60</sup> (ms)	46.1	145.5	21	7.85	0.00

Table 4.2: Mean durations of the total duration of vowel + plosive and of the duration of the plosives alone, and the statistical results. Speaker sp180, spontaneous speech.

Spontaneous speech	Long vowel + lenis plosive	Short vowel + fortis plosive	n	t	P
Total duration (ms)	147.6	183.4	23	2.36	0.01
Plosive duration (ms)	47.6	121.9	23	8.29	0.00

Table 4.3: Mean durations of the total duration of vowel + plosive and of the duration of the plosives alone, and the statistical results. Speaker sp127, spontaneous speech.

Once again, no statistically significant correlation can be found between vowel duration and consonant duration ( $r = 0.01$ ,  $p = 0.93$  for speaker sp180 and  $r = -0.29$ ,  $p = 0.09$  for speaker sp127).

These results reveal several things. Firstly, the Central Bavarian roots of Standard Austrian German become apparent. Secondly, since the adoption of non-Bavarian combinations (especially the “long vowel + fortis plosive” combination) did not lead to isochrony, isochrony is a consequence of the combination type and rather than an indication of an underlying temporal organisation. The results point to an independent treatment of vowel and plosive durations. The logical conclusion is that the speech chain is modelled phoneme by phoneme, as has already been pointed out by many researchers (e.g. Wood 1996, 1997, Lindblom & Sussman 2002). Thirdly, and perhaps most importantly, the observed temporal organisation of some vowel + plosive combinations is restricted to the sentence reading task. This might indicate that rhythmic patterning depends strongly on the speaking style (as becomes apparent anyway from a comparison of e.g. poetical recitations and spontaneous speech).

<sup>60</sup> It should be noted that in Standard Austrian German, the “fortis” and “lenis” plosives are not differentiated according to VOT. VOT can be present in those plosives that are generally labeled as “fortis”, whereas those plosives which are labeled as “lenis” can be deleted or spirantized or articulated as approximants (but see Moosmüller & Ringen 2004 for a detailed discussion).

Therefore, an assumption that Standard Austrian German would level out the differences in vowel length by adjusting the length of the following plosive is not justifiable. Consequently, for Standard Austrian German, a temporal analysis based on a durational compensation of vowel + consonant can be discarded.

#### **4.2.4. Vowel duration in Standard Austrian German**

It has been argued in 4.2.1 that vowels in Standard German contrast in quality, with duration being an accompanying feature reflecting the historical development. I.e. the [+tense] vowels are long and the [–tense] vowels are short. Antoniadis & Strube (1984), in their analysis of logatomes read by North German speakers, found out that long vowels were at least twice as long as their short counterparts. Iivonen (1987b) arrived at similar results for Standard Austrian German. He compared vowel duration in read monosyllabics between speakers of Austrian Standard (mostly from Vienna) and speakers of East Central German (from Halle) and found out that the duration of all vowels was 32% higher for the Austrian speakers than for the German speakers. However, the durational ratio between ascribed long/tense vs. short/lax vowels was about the same between the two regions, namely 2.2 (Iivonen 1987b: 326). This ratio suggests a high correlation between the feature [±tense] and duration, with the [+tense] vowels being at least twice as long as the [–tense] vowels. It has to be noted that the results are based on reading a list of monosyllabic words.

Duration, however, is composed of a certain number of periods and consequently contains information about fundamental frequency. From a psychoacoustic point of view, at least 8 periods are necessary for the perception of pitch<sup>61</sup>. In order to guarantee pitch perception in short vowels, the number of periods should be adapted to these perceptual needs. Such adjustment can, but need not, be accompanied by a rise in F0, since – from a long to a short vowel – only the number of periods and not necessarily

---

<sup>61</sup> I thank Werner Deutsch for pointing out this highly relevant connexion.

the duration of the periods undergo changes in the first place. Therefore, it is not surprising that absolutely no correlation can be observed between vowel duration and F0, whereas a high correlation can be found between vowel duration and the number of periods. This result holds for logatome reading task, the sentence reading task, and spontaneous speech. Table 4.4 gives the correlation coefficient  $r$  for all speakers:

$r$	Logatomes NoP/ms	Reading NoP/ms	Spontaneous NoP/ms	Logatomes ms/F0	Reading ms/F0	Spontaneous ms/F0
Sp012	<b>0.96</b>	<b>0.87</b>	<b>0.94</b>	<b>0.16</b>	0.02	0.07
Sp180	<b>0.94</b>	<b>0.84</b>	<b>0.89</b>	0.30	<b>0.23</b>	0.11
Sp082	-	<b>0.89</b>	<b>0.83</b>	-	<b>0.22</b>	0.11
Sp129	-	<b>0.91</b>	<b>0.92</b>	-	<b>0.17</b>	<b>0.22</b>
Sp126	-	<b>0.92</b>	<b>0.96</b>	-	<b>0.15</b>	0.12
Sp127	-	<b>0.85</b>	<b>0.96</b>	-	<b>0.25</b>	<b>0.17</b>

Table 4.5: Correlation coefficient  $r$  for NoP/ms and F0/ms for three speaking tasks, stressed vowels, all speakers. Statistically significant results ( $p < 0.05$ ) are in bold.

The number of periods (henceforth NoP) does not contain any information about F0 and can therefore be taken as a pure durational measure<sup>62</sup>. Given the high correlation between duration (in ms) and the number of periods, it can be concluded that the longer a segment, the more periods it contains. Since fundamental frequency is hardly affected by a truncation of periods, the number of periods renders more exact results than traditional duration measurements.

The results on NoP display a high dependency on the speaking task. In the logatome reading task, speaker sp180 discerns long and short vowels, i.e. vowels termed as “tense” have a higher amount of NoPs than the vowels labeled “lax” (27 vs. 14,  $n = 313$ ,  $t = 22.58$ ,  $p = 0.00$ ). The vowel /a/ is differentiated for NoP as well (27 vs. 14,  $n = 47$ ,  $t = 10.74$ ,  $p = 0.00$ ). Duration, therefore, coincides with tenseness. Figure 4.6 shows the results of a cluster analysis for speaker sp180, which groups long/tense vowels and short/lax vowels into two clusters:

<sup>62</sup> In the chapters 3.2.2 and 3.2.3, NoPs could not be calculated, because no periodic signal is available in the closure phase of unvoiced plosives.

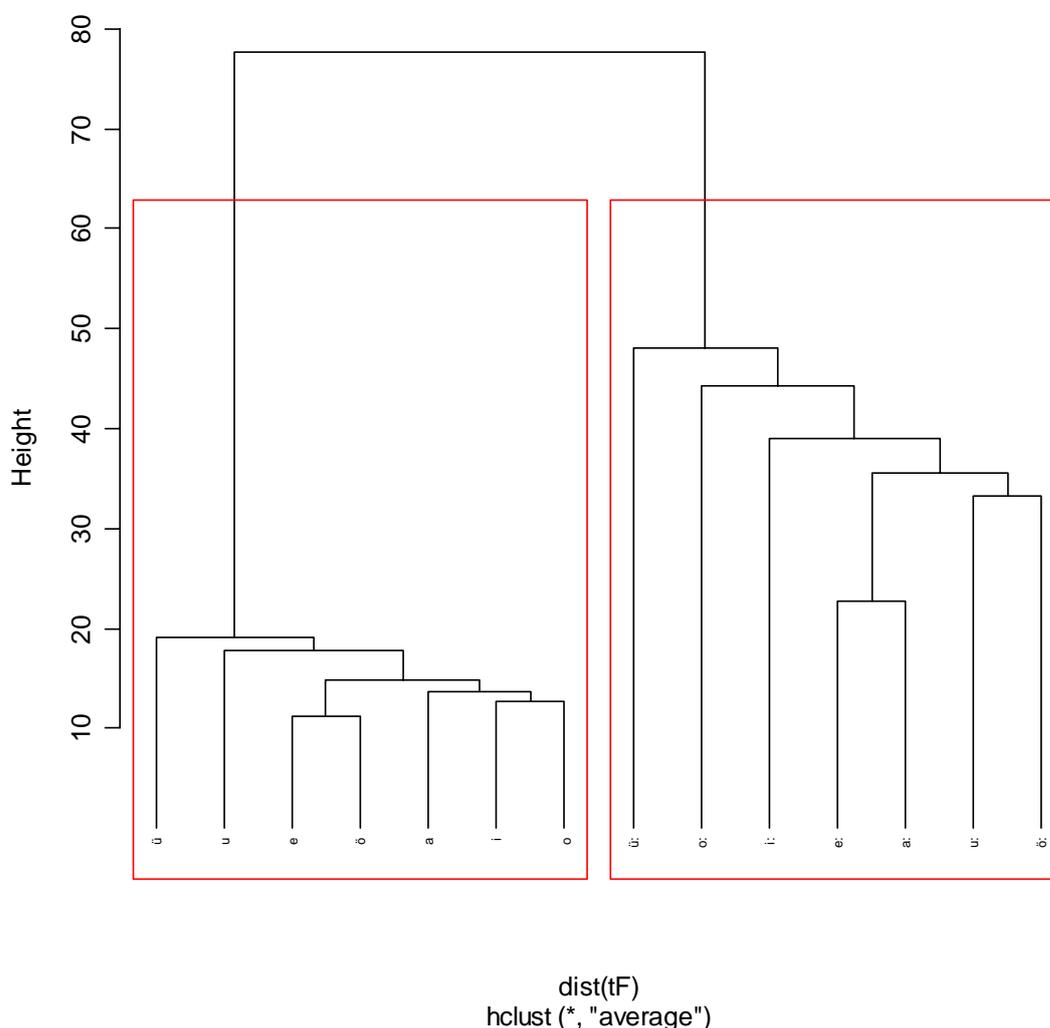


Figure 4.6: Cluster analysis for number of periods of long/tense and short/lax vowels in stressed positions, speaker sp180, logatome reading task. All short/lax vowels are grouped in the left box, all long/tense vowels in the right box.

It can be seen from the lengths of the vertical lines (indicating the distances to the other cluster members<sup>63</sup>) in Figure 4.6, that the distances among the long/tense vowels are much higher than the distances among the short/lax vowels. Again, for speaker sp012; vowels labeled “tense” have a higher amount of NoP than vowels labeled “lax” (18 vs. 9,  $n = 273$ ,  $t = 36.18$ ,  $p = 0.00$ ), the vowel /a/ is differentiated for NoP as well (18 vs. 9,  $n = 36$ ,  $t = 17.24$ ,  $p = 0.00$ ). Figure 4.7 shows the cluster analysis for speaker sp012.

<sup>63</sup> The distances, of course, contain no information about NoP.

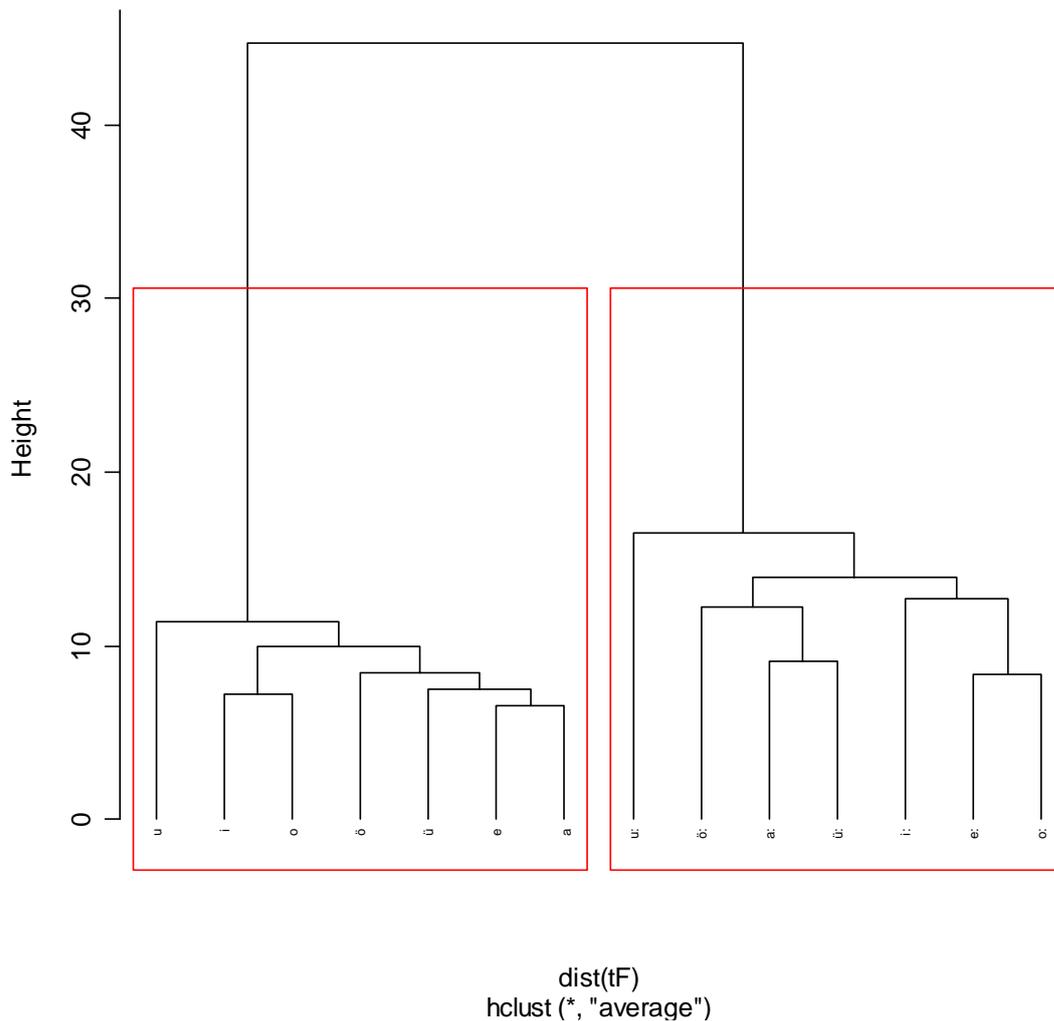


Figure 4.7: Cluster analysis for number of periods of long/tense and short/lax vowels in stressed positions, logatome reading task, speaker sp012.

For the other two tasks, the sentence reading task and the spontaneous speech task, the discriminatory ability of duration decreases step by step, till it finally stops fulfilling this function in unstressed positions (see Table 4.5 and 4.6: the results of the logatome reading task are repeated for better comparison):

Speaker sp180	NoP V “tense”	NoP V “lax”	n	t	p
Logatome reading task	27	14	313	22.58	0.00
Sentence reading task stressed	22	15	341	9.19	0.00
Spontaneous speech stressed	16	13	185	3.32	0.00
Sentence reading task unstressed	11	12	446	-2.82	0.00
Spontaneous speech unstressed	9	10	349	-1.23	0.11

Table 4.5: Mean number of periods for tense and lax vowels, and statistical results for 3 speaking tasks and 2 prosodic positions, speaker sp180.

Speaker sp012	NoP V “tense”	NoP V “lax”	n	t	p
Logatome reading task	18	9	273	36.18	0.00
Sentence reading task stressed	14	8	375	12.20	0.00
Spontaneous speech stressed	9	6	83	3.69	0.00
Sentence reading task unstressed	6	6	441	-3.52	0.00
Spontaneous speech unstressed	5	5	96	-0.17	0.43

Table 4.6: Mean number of periods for tense and lax vowels, and statistical results for 3 speaking tasks and 2 prosodic positions, speaker sp012.

These results also hold for the remaining speakers. It should be noted, that in the unstressed position, the situation is reversed: the “tense” vowels are somewhat shorter than the “lax” vowels, as is indicated by the negative t-value. This inversion is sometimes statistically significant, sometimes not. But the statistical significance is not important in this case, since a parameter which is reversed under certain conditions cannot make up a meaningful feature. As a summary, the NoP ratios for each person under each condition can be read from Table 4.7.

NoP Ratio	Logatomes	Reading stressed	Spontaneous unstressed	Reading unstressed	Spontaneous unstressed
Speaker012	1.94	1.67	1.48	1.13	1.02
Speaker180	1.99	1.47	1.25	1.13	1.06
Speaker126	-	1.50	1.14	1.22	1.18
Speaker127	-	1.51	1.43	1.04	1.06
Speaker082	-	1.39	1.19	1.19	1.02
Speaker129	-	1.80	1.32	1.09	1.06

Table 4.7: NoP ratio of tense/lax vowels grouped over all speakers according to speaking tasks and prosodic position.

It becomes evident from Table 4.7 that the NoP ratio, and consequently the NoP, are a function of speaking tasks and prosodic position, i.e. the more informal the speaking task and the weaker the prosodic position, the less differences can be observed between tense/long and lax/short vowels with respect to NoP<sup>64</sup>. In reading logatomes, tense/long vowels are nearly twice as long as their lax/short counterparts. The results presented in Iivonen (1987b) give a still higher ratio, namely 2.2. Given the dependence of the ratio on speech style and prosodic position, the higher ratio observable in Iivonen's study is the result of a still more formal speech style: a list of isolated monosyllabics vs. bisyllabic logatomes in carrier sentences<sup>65</sup>.

It has been observed in many languages, including Standard German (Antoniadis & Strube 1984, Strange & Bohn 1998) that duration depends on vowel height and is considered to be a phonetic universal (Maddieson 1997). However, this could not be observed in Standard Austrian German. In the two reading tasks, one-way ANOVA rendered statistically significant differences between the vowels ("tense" and "lax" vowels treated separately), in spontaneous speech. Only speaker sp127 and speaker sp180 show statistically significant differences for both vowel groups, and speaker sp126 only for the "lax" vowels. The crucial point, however, is that the /a/ vowels do not expose the highest number of periods. Exactly which vowel shows the highest number of periods, differs from speaker to speaker (/o/ and /a/ for speaker sp082, /y/ and /u/ for speaker sp129, /u/ and /a/ for speaker sp180, /y/ and /œ/ for speaker sp012, /ɑ/ and /a/ for speaker sp126, /y/ and /a/ for speaker sp127 in the sentence reading task, see Table 4.8). Though statistically significant, no meaningful order can be worked out. Table 4.8. gives an overview of the mean number of periods in the sentence reading task for all speakers:

<sup>64</sup> The same result was achieved by Strange & Bohn (1998) for Northern German.

<sup>65</sup> Eberhard Zwirner and his colleagues, who performed large-scaled regional investigations on vowel durations, found out that in the east and southeast of the German language area the durational ratio was only 1.1, whereas the ratio continuously increases to the west and the north and finally arrives at 2 in Bremen (Zwirner 1962, cited after Sendlmeier 1985).

Number of periods	Sp082	Sp129	Sp180	Sp012	Sp126	Sp127
/i/	18	18	18	11	12	14
/y/	23	30	22	21	16	24
/e/	21	20	22	14	15	17
/ø/	23	23	24	16	15	20
/ɑ/	20	18	22	12	17	17
/u/	21	19	28	15	16	18
/o/	24	21	25	17	17	19
/ɪ/	12	10	15	7	8	9
/ʏ/	15	10	14	8	9	11
/ɛ/	16	11	13	9	11	11
/œ/	17	13	17	9	11	12
/a/	18	12	18	9	12	13
/ʊ/	15	13	16	9	7	11
/ɔ/	13	9	14	8	10	11

Table 4.8: Mean NoP for all speakers, broken down for vowel type, sentence reading task.

The temporal patterns (in NoPs) observed for Standard Austrian German vowels continue a development that is already evident in Standard German. In Standard Austrian German, the contrastive power of NoP monotonously decreases with decreasing formality and prosodic strength. Moreover, no durational differences between high and low vowels can be observed; NoP ratios of the low vowels pairs are not bigger than the ratios of the high vowel pairs, i.e. the contrastive power of NoP is not any better for low vowels than for high vowels. From this it follows that a separate feature [ $\pm$  long] to discriminate the low vowels is not justified. It can be concluded that duration has no contrastive function in the Standard Austrian German vowel system.

#### 4.2.5. Duration and Laxness

In 4.2.4 it has been argued that duration is not a distinctive feature in Standard Austrian German. The historical analysis presented in Ronneberger-Sibold (1999) supports this analysis: since isochrony has been given up in Standard German, the quality of the former short vowels changed, i.e. the short “tense” vowels became “lax”.

From these historical facts it is often assumed that “lax” vowels and vowel shortening have something in common. Fischer-Jørgensen (1990) argues for such analysis:

“...the fact that only the short vowels have become lax in these Germanic languages also supports an interpretation of the historical development of these vowels as a reduction.” (Fischer-Jørgensen 1990:108).

Such an interpretation would imply that a) lax vowels become tense when lengthened and b) tense vowels become lax when shortened.

Arguments against mingling vowel quality with durational phenomena have already been propounded in Wood (1982: 157, see also Pulleyblank 2003). Wood’s basic counterargument is the fact that long lax vowels (/ɛ:/ and /ɔ:/ as e.g. in the Viennese Dialect, see Dressler & Wodak 1982, Moosmüller 1987) and short tense vowels exist (again e.g. in the Austrian dialects, see Dressler & Wodak 1982, Moosmüller 1987). Hoole et al. (1994) found out that, in dependence on speech tempo, it is especially the nucleus of tense vowels that is substantially contracted in fast speech styles, whereas the nucleus of lax vowels is hardly affected at all (52.3% vs. 12.5% shortening of the respective nuclei). However, vowel opposition is neither neutralized across speech tempi (Mooshammer 1998), nor across stress modi (Jessen 1993, Jessen et al. 1995). Since duration is not contrastive in Standard Austrian German, i.e. both short and long vowels are found in each category, it can easily be tested whether vowel quality changes in dependence on vowel duration.

Vowel quality has, for this test, been defined as a change in the mean<sup>66</sup> value of one of the first three formant frequencies. Correlation coefficients  $r$  have been calculated for NoP and one of the first three formants. The one-way ANOVA showed no significant differences between the results of the six speakers. Therefore, only the results of speaker sp012 are presented in Table 4.9 to provide the overall results.

---

<sup>66</sup> The mean value of each formant frequency is the result of the dynamic change over time of the vowel under consideration, i.e. formant transitions enter into the mean value.

Vowel category	Variability Coefficient NoP	r: NoP/F1	r: NoP/F2	r: NoP/F3
/ø/	34.39	0,26	0.43	<b>0.88</b>
/œ/	21,39	0.57	0.22	0.19
/y/	48,97	0.26	0.17	0.48
/ʏ/	34.19	0.52	<b>0.76</b>	0.52
/u/	42.30	0.18	0.25	0.06
/ʊ/	33,29	0.02	0.28	0.34
/o/	28.29	0.36	0.11	0.02
/ɔ/	42.08	0.06	0.23	0.06
/i/	36.64	0.06	<b>0.45</b>	<b>0.56</b>
/ɪ/	31.59	<b>0.36</b>	0.16	0.17
/e/	33.81	0.27	0.31	0.10
/ɛ/	34.74	0.07	0.03	0.13
/a/	32,78	<b>0.49</b>	0.03	0.19
/ɑ/	29.72	<b>0.56</b>	0.10	<b>0.39</b>

Table 4.9: Variability coefficient of NoP and correlation coefficient r for NoP/F1, F2, and F3 for each vowel. Statistically significant results ( $p < 0.05$ ) are in bold. Speaker sp012, sentence reading task, stressed position.

As becomes evident from Table 4.9, although the variability of NoP within each category is very high (see variability coefficient), the quality of the vowel does not change in dependence of NoP. This resistance to durational changes also becomes clear from the within category variability of the first three formants (see Table 4.10).

It can be seen from Table 4.10 that the variability coefficient for each formant is substantially smaller than the variability coefficient for NoP. From these results it can be concluded that in the stressed position, vowel quality is preserved within each vowel category, despite substantial variability in vowel duration. Moreover, it can be concluded that – contrary to the assumption proposed in Jakobson & Halle (1961/2002<sup>3</sup>) – vowels labeled as “lax” are not mere attenuated variants of their tense counterparts, but autonomous phonemes with separate articulatory settings.

Vowel category	Variability Coefficient NoP	Variability Coefficient F1	Variability Coefficient F2	Variability Coefficient F3
/ø/	28,71	8,665	5,481	3,877
/œ/	18,94	5,544	5,278	3,368
/y/	38,45	9,259	6,331	5,067
/ʏ/	30,18	10,548	9,520	5,197
/u/	31,44	7,689	14,375	7,169
/ʊ/	37,75	7,476	16,541	5,887
/o/	36,73	7,418	10,307	7,387
/ɔ/	28,04	7,271	10,188	6,662
/i/	38,04	7,874	4,870	8,512
/ɪ/	32,89	8,134	4,815	6,980
/e/	33,27	7,120	4,411	7,716
/ɛ/	30,83	6,475	5,513	3,758
/a/	31,33	9,982	7,788	7,327
/ɑ/	29,60	9,181	7,064	4,335

Table 4.10: Mean variability coefficients of NoP, F1, F2, and F3 for each vowel category, pooled over six speakers.

### 4.3. Articulatory investigations on vowels

#### 4.3.1 Articulatory settings for the opposition traditionally termed [ $\pm$ tense]

As has been already stated, it is not quite clear which articulatory settings are responsible for the difference in vowel quality for the pairs /i/ – /ɪ/, /e/ – /ɛ/, /y/ – /ʏ/, /ø/ – /œ/, /u/ – /ʊ/, /o/ – /ɔ/, and /ɑ/ – /a/. The vocal tract configuration necessary for the production of a given vowel contrast can be realized in many ways and by many different combinations of individual articulators. The feature [ $\pm$ tense] is probably the most disputed of all features discriminating vowels, since many different articulatory adjustments might cause the above mentioned oppositions and the articulatory adjustments involved are language specific. For example, the feature [ $\pm$ ATR], generally agreed to be of relevance in several African languages, which yields similar oppositions to those caused by [ $\pm$ tense] in Germanic languages, shows different articulatory ad-

adjustments with consecutively different acoustic outputs within the languages described as [ $\pm$ ATR]. The main physiological correlate of [+ATR] languages is either a widening of the pharyngeal cavity as a result of tongue root advancement or a reduction of the pharyngeal cavity by tongue root retraction, resulting in a lower F1 of the [+ATR] vowels as opposed to the [-ATR] vowels. These articulatory adjustments have been described for Igbo (Ladefoged 1964), Akan (Lindau 1979, 1987, Tiede 1996), Degema (Fulop et al. 1998), as examples of Niger – Congo languages, and for DhoLuo, Shilluk, Dinka (Jacobson 1978, 1980) and Maa (Guion et al. 2004) as examples of Nilotic languages. Additionally, in Akan, tongue root advancement is accompanied by a lowering of the larynx (Lindau 1974, cited in Guion et al. 2004, Tiede 1996). Tongue root advancement might be followed by a tension in the vocal folds which results in higher breathiness and shows up as a higher amount of energy in the higher frequency regions. However, these accompanying settings have only been observed for Akan (Hess 1992), Degema and Maa (in Maa with no statistical significance).

The feature [ $\pm$ ATR] has also been used to describe the “tense/lax” opposition in English (Halle & Stevens 1969)<sup>67</sup>. However, a cineradiographic study conducted by Ladefoged et al. (1972) showed that tongue root advancement is just a further complementary strategy to express the tense/lax opposition in English. This finding is corroborated by Jackson (1988). In his study on various languages he proved that English does not show a separate control of the tongue root for the tense/lax contrast. Tiede (1996) compared Akan and English and detected substantial differences in the articulatory patterning between the two languages when producing the desired distinctions. He concludes that

Akan and English show different patterning of axial data at measured levels below the epiglottis. With one exception (area measured at the three lowest levels of /e/), the area, width, and depth measurements obtained for Akan show consistently larger values for expanded (+ATR) variants at all measured levels, above and below the epiglottis. But while the English data also show consistently larger values above the epiglottal pivot, at levels

---

<sup>67</sup> Following Halle & Stevens (1969), Wiese (1996) adopts the feature [ $\pm$ ATR] for the vowels of “Modern Standard German” (1996: 20f).

below that point differences between tense and lax variants are inconsistent in sign and considerably smaller in magnitude. (Tiede 1996: 415f)

And he continues:

The English sagittal measurements showing smaller differences in magnitude for both tongue root advancement and laryngeal lowering than Akan, and greater differences in tongue dorsum height, suggest a relatively more significant role for tongue height in maintaining the English contrast. (Tiede 1996: 418)

Tiede (1996) also elaborates on the relationship between dorsum height and tongue root advancement, two adjustments that may be intrinsically linked. Whereas Akan, however, controls the muscle responsible for a concomitant change in tongue height for [+ATR] vowels, tongue height is accompanied with tongue root advancement in English. Therefore, it can be inferred that the primary feature used to express the tense/lax distinction in English is tongue height with facultative accompanying tongue root advancement, whereas in Akan the distinction is expressed primarily by tongue root advancement. The accompanying raise of the tongue dorsum is suppressed.

Ladefoged and Maddieson (1996) come to similar conclusions:

In Igbo and Akan the tongue height is not correlated with the tongue root position. In English the position of the tongue root is correlated with the tongue height.....We conclude that the advancement of the tongue root is a separable tongue gesture in languages such as Igbo and Akan. In Germanic languages, however, it is simply one of the concomitants of vowel height. (Ladefoged & Maddieson 1996: 303f)

Recently, following up the hypothesis proposed by Halle & Stevens (1969), Slifka (2003) tested F1 slope (indicating breathiness) and point in time of the energy peak for tense/lax vowel pairs in English. Although 88,9% of all tense vowels exposed a falling slope and 91,7% of all lax vowels exposed a rising slope, indicating more breathiness in the tense vowels, her results are not consistent across the speakers tested and ultimately fail to classify vowels correctly. It has to be noted, furthermore, that breathiness is not only a facultative consequence of advanced tongue root, but also of tongue fronting and tongue heightening. Therefore, the feature ATR cannot be inferred from increased breathiness. Consequently, the feature [ $\pm$  ATR] is not responsible for the opposition termed [ $\pm$  tense] in the Germanic languages.

Wood (1975b, 1979) found constriction degree contrastive for the tense/lax opposition in English and Egyptian Arabic. The degree of constriction is narrower for tense vowels in all pairs except [o, ɔ]. In this pair, the constriction is wider for [o] than for [ɔ], “although both ranges virtually overlap” (1975b: 112 ff). In their investigation of German vowels, Pouplier et al. (2004) found no inverse relationship of tongue-palate distances of the vowel pair [o] – [ɔ], i.e. the tongue-palate distance for [ɔ] was greater than for [o]. In the investigation of Pouplier et al. (2004), tongue-palate distance was neutralised for [i] – [ɪ] in all subjects and for [u] – [ʊ] in two subjects. For German, Hoole & Mooshammer (2002) confirmed Wood’s results on the front vowels: /e/ exposes a higher tongue position than /ɪ/. The results presented in Pouplier et al. (2004) show intersubject variability: two subjects expose no differences, one subject has a higher tongue-palate distance for /e/. For the front unrounded vowels Wood (1982) concludes

“that the tongue is higher and more bunched relative to the mandible for the tense vowels [i, e] and lower, flatter and bulging further into the pharynx for the lax vowels [ɪ, ɛ]. This difference is performed with the mandible raised (for close [i, ɪ]) and lowered (for open [e, ɛ])”. (Wood 1982: 140f).

The results on the front unrounded vowels presented by Wood (1982) and corroborated by Fischer-Jørgensen (1990) and Valaczkai (1998)<sup>68</sup> clearly set apart two features which are often used synonymously, namely the degree of constriction<sup>69</sup> (tongue height) and the degree of jaw opening, i.e. tongue height is not accompanied by a concomitant adjustment of jaw position<sup>70</sup>. It follows that the “e-vowels” and “i-vowels” are grouped by jaw opening – the opening of the e-vowels is larger than 8 – 9 mm, whereas the

---

<sup>68</sup> Although in Valaczkai (1998), the degree of lip opening is greater for the respective “lax” vowel.

<sup>69</sup> For the front vowels, constriction degree is equivalent to tongue-palate distance.

<sup>70</sup> Sievers (1901:100), who introduced the features “tense” and “lax”, also clearly distinguishes this pair from the pair “open” and “close”: “Man hüte sich auch davor, die Begriffe ‘gespannt’ (oder ‘eng’) und ‘ungespannt’ (oder ‘weit’) mit denen zu verwechseln, welche die althergebrachten Ausdrücke ‘geschlossen’ und ‘offen’ bezeichnen sollen. Diese letzteren wollen nur aussagen, dass ein Vocal geringere oder grössere Mundweite habe als ein anderer, aber ohne alle Rücksicht auf die Verschiedenheit der Articulationsweise, welche die Differenzen der Mundweite im einzelnen Fall hervorruft...”

opening for the “i-vowels” is smaller than 8 – 9 mm (Wood 1982: 46). The “tense” vs. “lax” vowels, on the other hand, are grouped by constriction degree – for /i, e/ the tongue is higher than for /ɪ, ɛ/.

Unfortunately, results cannot be grouped as nicely for the back rounded vowels. First of all, degree of jaw opening does not group the “o-vowels” against the “u-vowels”. /u/ has nearly the same degree of openness as /o/, and, consequently, both /u, o/ are closer than /ʊ/, which again is closer than /ɔ/. The front rounded vowels do not differ too much with respect to degree of jaw opening, but the tendency is the same as for the back rounded vowels (Valaczkai 1998). This renders the following pattern:

	<b>Close</b>	<b>Open</b>
<b>Front unrounded</b>	i, ɪ	e, ɛ
<b>Back rounded</b>	u, o	ʊ, ɔ

The lip configuration of the “tense” back rounded vowels is more protruded than for their “lax” cognates (Wood 1982 :145). Consequently, the degree of opening is narrower for /u, o/ than for /ʊ, ɔ/. For the front rounded vowels, only a moderate lip protrusion is favourable in order to ensure F3 and F2 at close quarters (Wood 1986). The moderate lip protrusion in front rounded vowels explains the fact that there are hardly any differences with respect to jaw opening between the tense and the lax front rounded vowels.

Lip protrusion is accompanied by larynx depression, in order to adjust the distance from the glottal source to the constriction (Wood 1979: 33). Therefore, the degree of larynx lowering depends on the degree of lip protrusion; i.e. for the more protruded “tense” vowels, the larynx is lowered by approximately 10 mm, whereas it is only lowered by approximately 5 mm for the “lax” vowels, which expose a minor protrusion (Wood 1979).

Pouplier et al. (2004) found initial genioglossus posterior (GGP) compression for all “tense” vowels<sup>71</sup>, whereas for the “lax” vowels, immediate GGP expansion could be observed. They conclude:

“The estimation of genioglossus activity revealed the most consistent difference pattern across vowels [...]. Genioglossus posterior showed a consistent difference in that for the lax vowel, there was an immediate expansion movement, while the tense vowel either showed no change or compression, notably without corresponding expansion in the tongue tip [...]. The consistent differences that could be observed for the vowels across the entire utterance are consistent with the hypothesis that the tenseness opposition is not so much realized at the maximum constriction but rather lies in the entire motion sequence into and out of the vowel.” (Pouplier et al. 2004: 53)

This finding would imply that the main difference between “tense” and “lax” vowels lies in the onset of the vowel. It has been stated above that Pouplier et al. (2004) also found differences in the tongue-palate distance, except for the fact that these differences were neutralized for the pairs [i] – [ɪ] and [u] – [ʊ]. For this reason, tongue-palate distance was discarded as a discriminatory parameter. However, from a phonological point of view, neutralization of these vowel pairs is possible. Therefore, it could in the same way be the case that GGP activity and tongue-palate distance form the articulatory correlate for the opposition [± tense], and that tongue-palate distance loses its discriminatory ability for the vowel pairs [i] – [ɪ] and [u] – [ʊ]. Consequently, for these pairs, differences should only be apparent in the onset of the vowel. However, according to the results presented in Pouplier et al. (2004), the pair [y] – [ʏ] does not follow this development. It has to be emphasized that tongue-palate distances heavily depend on the shape of the palate (Perkell 1997, Brunner et al. 2005). Therefore, the tongue-palate distances have to be individually adjusted in order to obtain the desired acoustic output. These individual differences in tongue-palate distances have also been observed by Pouplier et al. (2004).

To summarize, the following articulatory correlates can be obtained from the literature:

- [± ATR]: can be excluded for Germanic languages.

---

<sup>71</sup> Results on the pair /ø – œ/ are not reported.

- [± constricted]: refers to constriction degree and is most probably the best candidate to describe the so-called tense/lax opposition.
- [± open] differs for vowels with and without lip protrusion. /i/ and /ɪ/ are closed, whereas /e/ and /ɛ/ are open. For the vowels with lip protrusion, the opposition goes hand in hand with the tense/lax opposition.
- [± compressed] refers to GGP compression at the onset of the “tense” vowels.

This last feature would also represent a meaningful correlate. However, the opposition would be restricted to the onset of the vowel. Moreover, Pouplier et al. (2004) only tested the phonetic environment /gVm/. Since only the onset is affected by different GGP activity, other consonantal environments would have to be analysed in order to corroborate these results.

It can be concluded that, for the time being, a meaningful correlate for the feature [± tense] cannot be identified. The feature [± constricted] points to the degree of constriction and must not be replaced by [± tense]. In fact, the feature [± tense] is also misleading since it denotes that the speaker has to exert more effort in producing the [+ tense] vowels, leading ultimately to higher articulatory precision. Wood (1982) proved, however, that lax vowels are by no means articulated with less precision:

“Regarding precision, it is fascinating to watch a motion X-ray film and see the level of precision achieved for all vowels, tense and lax.” (Wood 1982: 177).

The results, as concerns articulatory precision for both tense and lax vowels, have been confirmed by Hoole & Mooshammer (2002) in their work on German vowels. Therefore, in order to avoid the misleading semantic implications, which have no articulatory basis, the feature [± tense] should be abandoned. The feature [± constricted] is – for Standard Austrian German – a far better articulatory description of the vowel pairs in question. Moreover, in cases where /a/ and /a/ are not neutralised, this feature also captures this opposition, since /a/ exposes a higher constriction in the pharynx than /a/ (Wood 1975b, Valaczkai 1998, Pouplier et al. 2004). However, the feature [± front],

which does not refer to constriction location in the case of the /a/ – vowels, is not apt to distinguish the /a/ – vowels in German, because the velar fricative is not palatalised after the traditionally called front /a/ in words like “Bach” (brook) or “lachen” (to laugh) (see also 4.4.5).

### 4.3.2 The features [± front] and [± high]

Wood (1979, 1982) isolated four contrastive constriction locations for vowels: “along the hard palate, along the soft palate, in the upper pharynx and in the lower pharynx” (1982: 43), but does not relate them to the tense/lax opposition. The front region can be further divided into a pre-palatal and a mid-palatal region. Whether the front region is further divided is language specific. Fant (1965) states that the Russian and Scandinavian [i]-vowels are pre-palatal, whereas the English [i] is articulated more towards the mid-palatal region (1965: 137). Wood (1979) found that his

“English and Arabic subjects had strikingly different constriction locations for palatal vowels... The English subject centred his constrictions midway along the hard palate about 35 mm behind the central incisors. The Arabic subject’s constrictions were more anterior, about 27 mm behind the central incisors.” (Wood 1979: 34)

Similar results are presented in Fant (2001). Modelling Swedish vowels, he spotted a front region located less than 4 cm from the teeth. The constrictions of the vowels [y:, ɥ:, i:] are located at 27, 28 and 31 mm from the teeth respectively. The constrictions of [e:, œ:, ø:] are located at 35, 37 and 42 mm from the teeth respectively (Fant 2001: 47). I.e. there is a pre-palatal location for [y:, ɥ:, i:], a mid-palatal location for [e:, œ:, ø:]. In Mooshammer (1998), who analyzed German vowels, horizontal tongue position played an important contrastive role as well. Horizontal tongue position, also termed “front raising” (Harshman et al. 1977), was also proved to discriminate vowels in Ningbo Chinese (Hu 2003) and Ndumbea, an Austronesian language (Gordon & Maddieson 1999).

As concerns the back vowels, the constriction locations spotted in Wood (1982), i.e. the soft palate for the “u-vowels”, the upper pharynx for the “o-vowels”, and the lower pharynx for the “a-vowels”, have been confirmed by Fant (2001).

For Standard Austrian German, the parameter “constriction location” allocates the individual constricted/unconstricted pairs a specific location<sup>72</sup>. The following constriction locations are assumed:

- pre-palatal for the /i, ɪ, y, ʏ/ – vowels
- mid-palatal for the /e, ε, ø, œ/ – vowels
- soft-palatal for the /u, ʊ/ – vowels
- upper pharyngeal for the /o, ɔ/ – vowels and
- lower pharyngeal for the /ɑ, a/ – vowels

In the light of this analysis, it is of no relevance whether constriction degree is bigger or smaller for /e/ with respect to /ɪ/ or /i/<sup>73</sup>. Constriction degree is not a gradual movement from narrow to wide, but exposes distinctive adjustments. Pouplier et al. (2004) found tongue-palate distance differences > 1 mm for two subjects (out of three) for the pair /i/ – /e/, and for only one subject for the pair /ɪ/ – /e/. Therefore, in German, /e/ is not to be discerned by tongue height from the /i/ – vowels, but by constriction location. The same holds for the /o/ – vowels vs. the /u/ – vowels. According to the results presented in Pouplier et al. (2004), tongue-palate distance is not bigger for /o/ than for the /u/ – vowels. The assumption that /e/ and /o/ are “mid” vowels is, therefore, not justified for Standard Austrian German.

For acoustic reasons, tongue-palate distance is bigger for the front rounded vowels than for their unrounded cognates (Wood 1986, Pouplier et al. 2004, also Chapter 4.4). This leads to four different tongue-palate distances in the two front regions respectively. The vowel /œ/, however, can be modelled both as a back and as a front vowel (Boë et al. 1992). For this vowel, “the vocal tract most resembles a cylindrical tube” (Boë et al. 1992: 35) and “the area functions reveal two approximately equal and symmetric

<sup>72</sup> This analysis holds only for languages which discern a pre-palatal and a mid-palatal location.

<sup>73</sup> See also Hu (2003) for a discussion on tongue height.

minima around  $X_{c.g}^{74} = 5$  cm and  $X_{c.g} = 11$  cm (1992: 36). For German, Hoole & Mooshammer (2002) describe the vowel /œ/ as a front vowel. Pouplier et al. (2004) do not present results on tongue-palate distances for /ø/, whilst for /œ/, tongue-palate distance difference of /ɛ/ vs. /œ/ exceeds 1 mm. Although the greater tongue-palate distance for the front rounded vowels has an important acoustic function (Wood 1986), it is nevertheless not useful for discriminating the front rounded vowels from their unrounded cognates by tongue height, because the respective primary discriminatory features are [± round] and [± constricted]<sup>75</sup>. The feature [± high] can, therefore, be discarded. The feature [± front] is, however, useful for giving a rough classification of vowels with a constriction location approximately < 4 cm from the lower incisors, and for vowels with a constriction location approximately > 6 cm from the lower incisors.

However, the feature [± front] is not able to discern the pre-palatal vowels and the mid-palatal vowels on the one hand, the lower pharyngeal vowels, the upper pharyngeal vowels, and the velar vowels on the other hand. This can be achieved by introducing three further features referring to constriction location: [± lower pharyngeal], [± velar], and [± pre-palatal]. [± lower pharyngeal] sets the velar and upper pharyngeal vowels apart from the lower pharyngeal vowels, and the [± pre-palatal] sets the pre-palatal and mid-palatal vowels apart.

From an articulatory point of view, the following features can be set up for Standard Austrian German:

- [±constricted] discerns /i, y, e, ø, o, u, a/ from /ɪ, ʏ, ɛ, œ, ɔ, ʊ, a/
- [±round] discerns /y, ʏ, ø, œ, o, ɔ, u, ʊ/ from /i, ɪ, e, ɛ, a, a/
- [±front] discerns /i, ɪ, y, ʏ, e, ɛ, ø, œ/ from /u, ʊ, o, ɔ, a, a/
- [±lower pharyngeal] discerns /ɑ, a/ from /i, ɪ, y, ʏ, e, ɛ, ø, œ, u, ʊ, o, ɔ/
- [±velar] discerns /u, ʊ/ from /i, ɪ, y, ʏ, e, ɛ, ø, œ, o, ɔ, a, a/
- [±pre-palatal] discerns /i, ɪ, y, ʏ/ from /e, ɛ, ø, œ, o, ɔ, u, ʊ, a, a/

<sup>74</sup>  $X_{c.g}$  = constriction coordinate from the glottis.

<sup>75</sup> [±constricted] is, of course, a tongue height parameter as well. However, this feature keeps vowel pairs apart. Tongue height, on the other hand, traditionally discerns the /a/ – vowels ([– high]), the /e, o/ – vowels ([+ high, – high]), and the /i, u/ – vowels ([+ high]).

Table 4.11 presents the feature matrix of the vowels of Standard Austrian German assumed from the articulatory investigations so far:

	/i/	/ɪ/	/y/	/ʏ/	/e/	/ɛ/	/ø/	/œ/	/a/	/ɑ/	/o/	/ɔ/	/u/	/ʊ/
constricted	+	–	+	–	+	–	+	–	+	–	+	–	+	–
round	–	–	+	+	–	–	+	+	–	–	+	+	+	+
front	+	+	+	+	+	+	+	+	–	–	–	–	–	–
lower pharyngeal	–	–	–	–	–	–	–	–	+	+	–	–	–	–
velar	–	–	–	–	–	–	–	–	–	–	–	–	+	+
pre-palatal	+	+	+	+	–	–	–	–	–	–	–	–	–	–

Table 4.11: Feature matrix of the vowels assumed for Standard Austrian German so far.

#### 4.4. Acoustic analysis of the vowels of Standard Austrian German.

It is well known that changes in the quality of a given vowel are dependent on the speaking task, the speech situation, or the prosodic position. This variation is consistent, planned, and, consequently, predictable to a high degree (see Chapter 5). Therefore, it makes no sense to test whether a given vowel maintains one and the same quality over diverse speaking tasks. It is, however, of relevance, whether, within a given speaking task, the speakers uphold the oppositions assumed in Chapter 4.3, or whether any neutralizations within a certain speaking task or a certain prosodic position (be it, e.g., either in a stressed position or in spontaneous speech) can be observed in the data. Such neutralizations, where they are consistent over all speakers, might point to a sound change in progress. One-tailed t-tests have been performed to test whether the assumed oppositions are preserved within each speaking task for a given vowel pair. Neutralization was defined in cases where a given vowel pair showed no statistically significant differences, or where it showed a significant difference that points, however, in the wrong direction, e.g. a higher F3 value for /ɪ/ as opposed to /i/. These latter cases are marked as bold and italic in the tables.

### 4.4.1. The pre-palatal vowels

The unrounded pre-palatal vowels are located in an acoustically unstable region. Probably for this reason, this region is only exploited in vowel systems with many front vowels. Since German discerns eight front vowels, it is not possible to locate them all in the mid-palatal region with five distinctive vowel heights. It is, in principle, possible to discern five vowel heights. However, tongue-palate distance differences depend on speaking tasks, situations, and prosodic positions as well (e.g., in weak prosodic positions, tongue-palate distance becomes greater). This stepwise increase of tongue-palate distance per vowel in dependence on speaking task and prosodic position would ultimately lead to a collapse of the feature tongue height, in cases where all front vowels were located in the same place.

As has been elaborated in Chapter 2, the pre-palatal location might cause a switch in cavity affiliation for F2 and F3, where constriction degree is small and constriction length is sufficiently long (approximately 5 cm). This vocal tract configuration causes F3 to raise substantially and approximate F4 (see Figure 4.8).

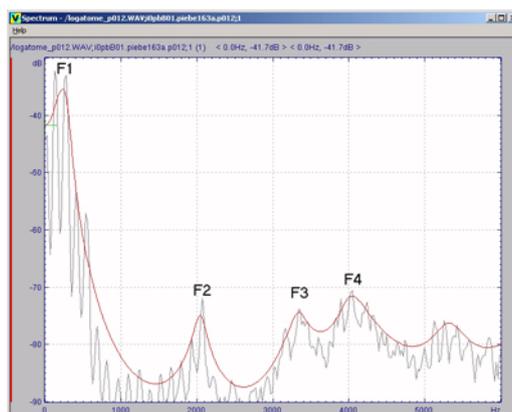


Figure 4.8: Averaged spectrum of the vowel /i/ taken from the logatome “piebe”, speaker sp012.

However, a configuration resulting in a spectrum like the one displayed in Figure 4.8 takes place at best in the most formal speech situations or speaking tasks. As soon as the

degree of formality decreases or prosodic position weakens, constriction length decreases, constriction degree widens, and F3 decreases again. The vowel /i/ is affected by this acoustic instability; the mean F3 value decreases by 12 % for reading sentences as compared to reading logatomes, whereas average F3 only decreases by 4.5 % for reading sentences as compared to spontaneous speech (see Table 4.14).

The instability of F3 of the vowel /i/ is most conspicuous in the intermediate speaking task of reading sentences, as becomes apparent from the density plot of all measured F3 values for the vowel /i/, broken down for three speaking tasks (Figure 4.9).

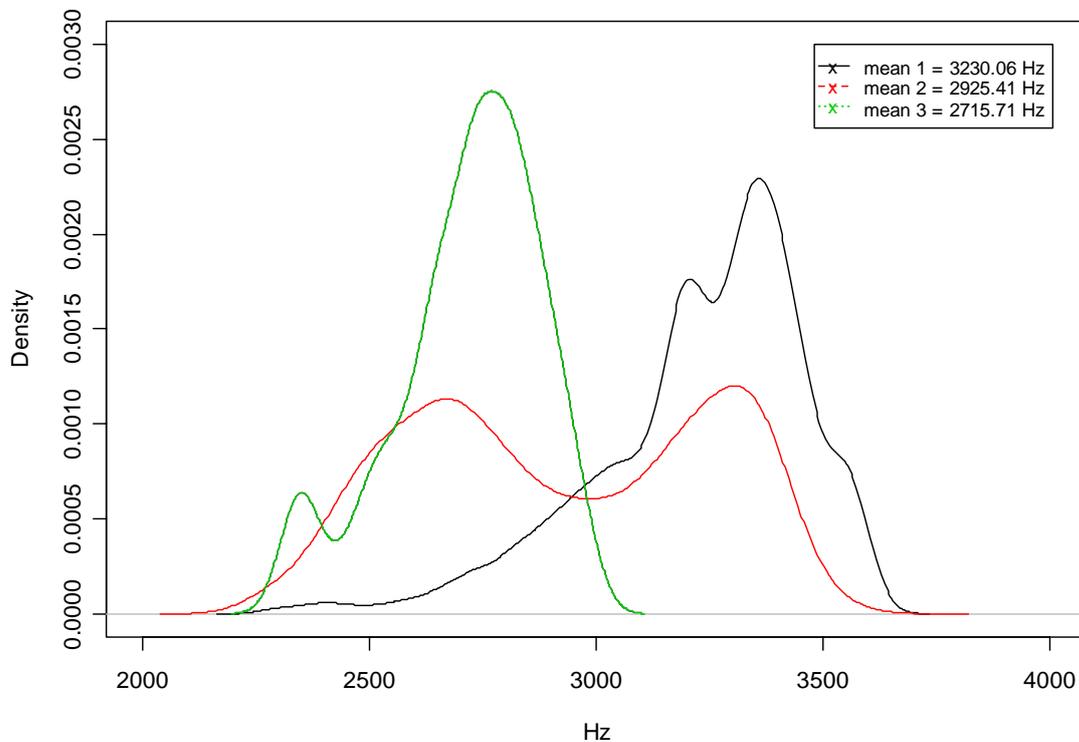


Figure 4.9: Density plot of all F3 values for the vowel /i/ in the stressed position, speaker sp012. Black line: logatome reading task, red line: sentence reading task, green line: spontaneous speech.

It can be seen from Figure 4.9 that F3 of the sentence reading task exhibits two evenly distributed peaks, one overlapping with the logatome reading task, indicating a front cavity affiliation for F3, the other overlapping with spontaneous speech, indicating a

back cavity affiliation of F3. Figure 4.9 also displays the stepwise change in vowel quality in dependence on the speaking task or degree of formality.

The [–constricted] unrounded pre-palatal vowel /ɪ/ is not affected by this problem, since the originally higher constriction degree prevents a switch in cavity affiliation. It can be seen from Figure 4.10 that F2, F3, and F4 are approximately evenly spaced.

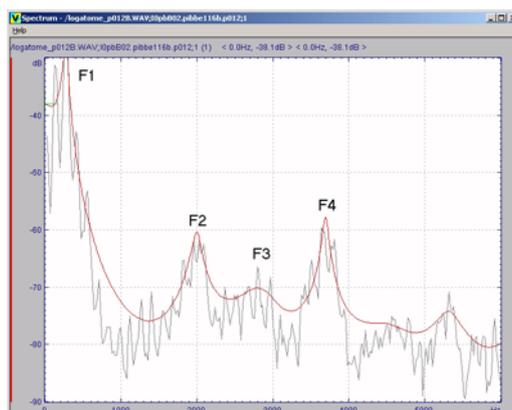


Figure 4.10: Average spectrum of the vowel /ɪ/ taken from the logatome “pibbe”, speaker sp012.

The primary difference between /i/ and /ɪ/ is tongue-palate distance, i.e. constriction degree. This parameter affects all formants; as constriction is widened, F1 increases<sup>76</sup>, and F2 and F3 decrease monotonously. Therefore, F1 has to be higher for /ɪ/ than for /i/, F2 has to be lower or equal<sup>77</sup>, and F3 has to be lower within each speaking task or prosodic position. However, Pouplier et al. (2004) pointed out that tongue-palate distances were neutralized between the pair /i/ – /ɪ/ for all three subjects. Tables 4.12 – 4.14 give the mean F1, F2, and F3 values for the pair /i/ – /ɪ/.

<sup>76</sup> However, the degree of lip aperture is the most obvious correlate of F1 (Hoole 1997).

<sup>77</sup> Depending on the cavity affiliation of F2.

F1	L	L	Rs	Rs	Rus	Rus.	Ss	Ss	Sus	Sus
Sp012	<b>269</b>	<b>281</b>	<b>284</b>	<b>298</b>	<b>317</b>	<b>328</b>	330	322	337	341
Sp180	<b>241</b>	<b>301</b>	<b>301</b>	<b>314</b>	320	319	306	314	341	346
Sp082	–	–	<b>379</b>	<b>392</b>	398	400	363	392	383	396
Sp129	–	–	<b>355</b>	<b>385</b>	<b>382</b>	<b>389</b>	<b>367</b>	<b>384</b>	<b>373</b>	<b>390</b>
Sp126	–	–	<b>282</b>	<b>296</b>	308	313	<b>302</b>	<b>313</b>	325	323
Sp127	–	–	<b>284</b>	<b>320</b>	<b>307</b>	<b>324</b>	311	310	320	327

Table 4.12: Mean F1 values of /i/ and /ɪ/ over all speaking tasks. Within each task, the value to the left represents the vowel /i/, the value to the right the vowel /ɪ/. Statistically significant differences within each pair ( $p < 0.05$ ) are marked in bold. Where  $t$ -values are positive, the pair is additionally in italics. L = Logatome reading task, Rs = Sentence reading task, stressed vowels, Rus = Sentence reading task, unstressed vowels, Ss = Spontaneous speech, stressed vowels, SuS = Spontaneous speech, unstressed vowels, Sp = Speaker.

F2	L	L	Rs	Rs	Rus	Rus.	Ss	Ss	Sus	Sus
Sp012	<b>2078</b>	<b>2032</b>	<b>2054</b>	<b>2022</b>	<i>1918</i>	<i>1955</i>	2079	1964	1758	1804
Sp180	<b>2533</b>	<b>2399</b>	<b>2489</b>	<b>2413</b>	<i>2256</i>	<i>2345</i>	2350	2345	2153	2201
Sp082	–	–	<b>2334</b>	<b>2217</b>	<i>2059</i>	<i>2128</i>	2179	2190	1907	1964
Sp129	–	–	2369	2348	2183	2235	<b>2379</b>	<b>2272</b>	2116	2155
Sp126	–	–	<b>2094</b>	<b>2030</b>	<i>1813</i>	<i>1895</i>	1985	1972	1744	1796
Sp127	–	–	<b>2110</b>	<b>2038</b>	1936	1910	2043	1952	<i>1705</i>	<i>1819</i>

Table 4.13: Mean F2 values of /i/ and /ɪ/ over all speaking tasks. Within each task, the value to the left represents the vowel /i/, the value to the right the vowel /ɪ/. Statistically significant differences within each pair ( $p < 0.05$ ) are marked in bold. Where  $t$ -values are negative, the pair is additionally in italics. Legend as in Table 4.12.

F3	L	L	Rs	Rs	Rus	Rus.	Ss	Ss	Sus	Sus
Sp012	<b>3230</b>	<b>2887</b>	<b>2828</b>	<b>2681</b>	<i>2508</i>	<i>2592</i>	2700	2622	2305	2389
Sp180	<b>3413</b>	<b>2979</b>	<b>3179</b>	<b>3065</b>	<i>2916</i>	<i>2968</i>	3042	3015	2781	2818
Sp082	–	–	<b>2893</b>	<b>2685</b>	<i>2509</i>	<i>2584</i>	2733	2707	2475	2539
Sp129	–	–	<b>3063</b>	<b>2921</b>	2782	2813	<b>2993</b>	<b>2807</b>	2656	2692
Sp126	–	–	<b>2697</b>	<b>2517</b>	<i>2351</i>	<i>2452</i>	2561	2587	<b>2270</b>	<b>2434</b>
Sp127	–	–	<b>2674</b>	<b>2543</b>	2439	2495	2516	2532	2403	2433

Table 4.14: Mean F3 values of /i/ and /ɪ/ over all speaking tasks. Within each task, the value to the left represents the vowel /i/, the value to the right the vowel /ɪ/. Statistically significant differences within each pair ( $p < 0.05$ ) are marked in bold. Where  $t$ -values are negative, the pair is additionally in italics. Legend as in Table 4.12.

It becomes apparent from Tables 4.12 – 4.14 that the opposition between the pair /i/ – /ɪ/ in stressed positions is fully maintained in the logatome reading task. It is also fully maintained in stressed positions in the task of reading sentences by five of the speakers. However, the opposition in stressed positions starts to be neutralized in spontaneous speech for all speakers except speaker 129. Figure 4.11 summarizes the results of the statistical analysis:

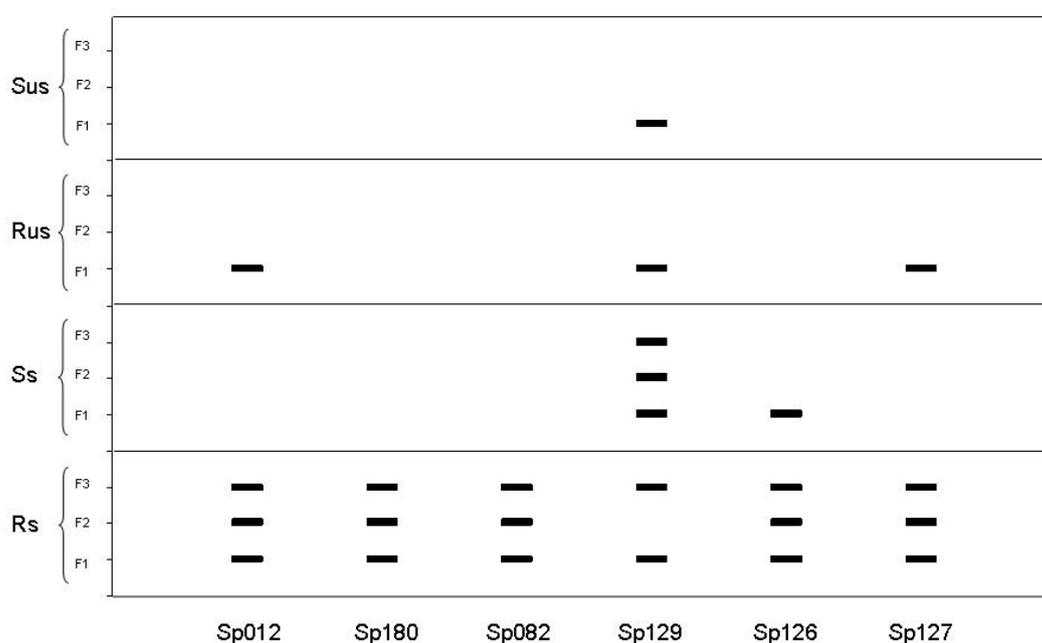


Figure 4.11: Results of the one-tailed t-tests for the vowel pair /i/ – /ɪ/. For each formant and each speaking task, statistically significant differences ( $p < 0.05$ ) are indicated by crossbeams. Where no differences occur for a given formant, the space is left blank. Legend as in Table 4.12.

Figure 4.11 shows that only two speakers distinguish the vowel pair in the stressed position in spontaneous speech; speaker sp129 distinguishes them by all three formants and speaker sp126 only by F1. In the unstressed positions, the vowel pair is, if at all, only discerned by F1. The cluster analysis presented for speaker sp012 illustrates the

differences between the speaking tasks. Figure 4.12 presents a cluster analysis for all /i/ and /ɪ/ vowels in stressed positions in the task of reading sentences:

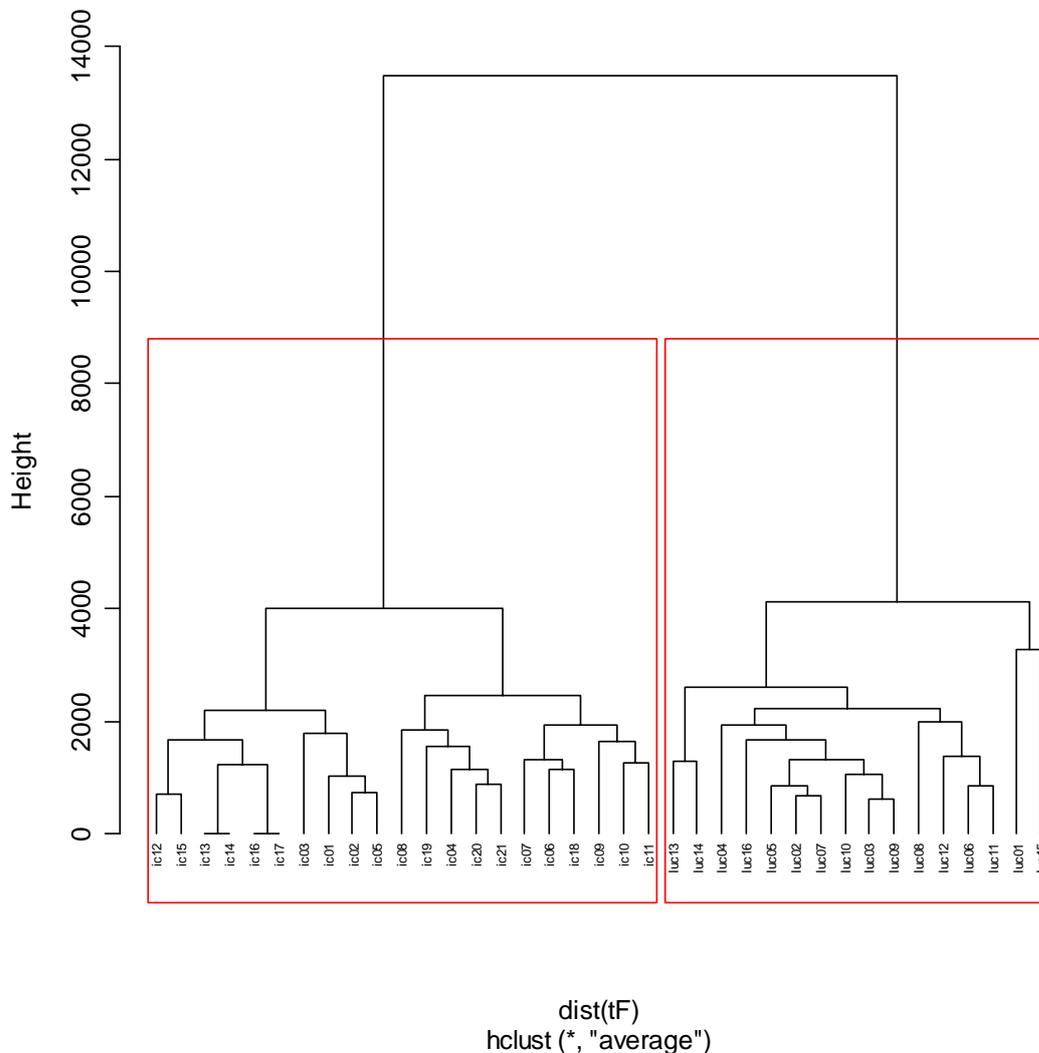


Figure 4.12: Cluster analysis of F1, F2, and F3 of all /i/ and /ɪ/ vowels in stressed positions. Sentence reading task. Speaker sp012. “ic” = /i/ [+constricted], “luc” = /ɪ/ [-constricted].

In Figure 4.12, the /i/ vowels are clearly kept apart from the /ɪ/ vowels. In unstressed positions in spontaneous speech, however, the opposition becomes totally blurred (Figure 4.13).

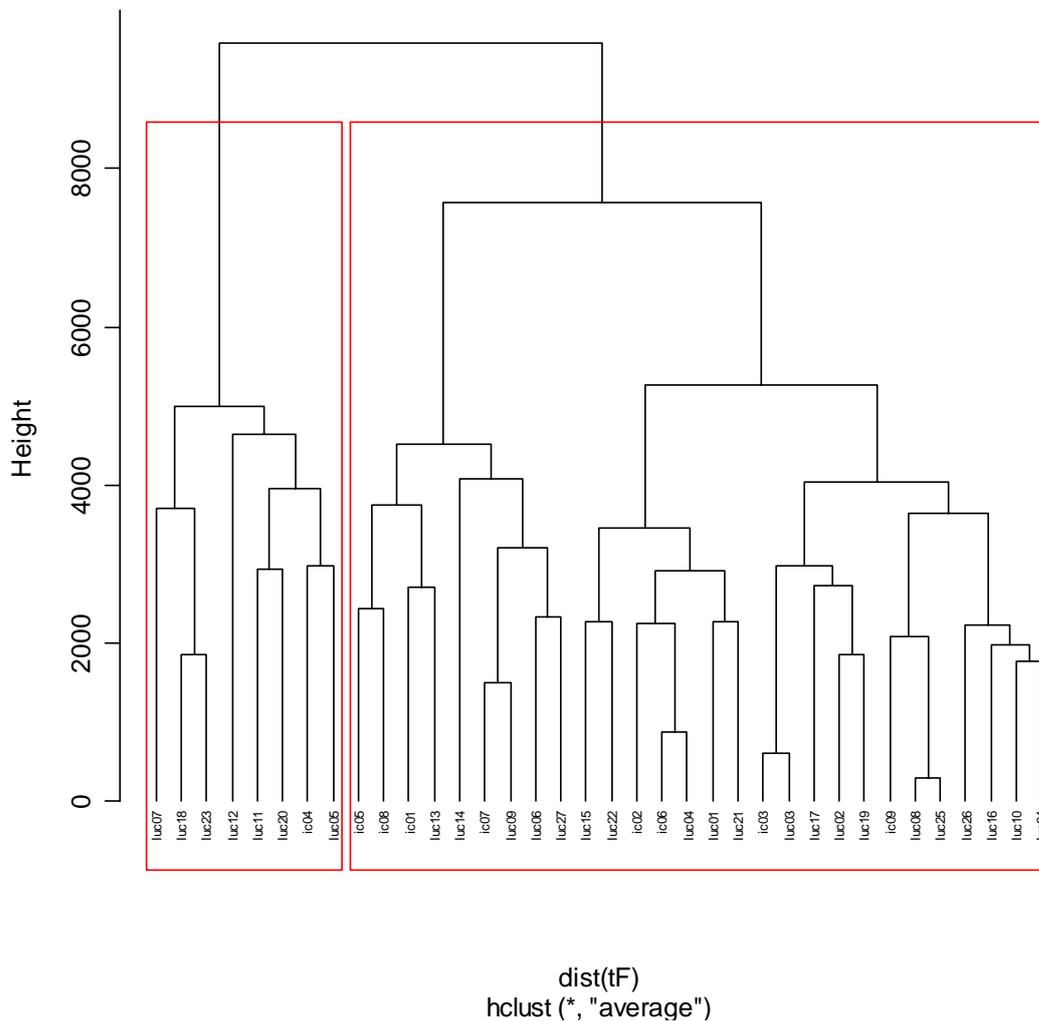


Figure 4.13: Cluster analysis of F1, F2, and F3 of all /i/ and /ɪ/ vowels in unstressed positions in spontaneous speech: speaker sp012. “ic” = /i/ [+constricted], “luc” = /ɪ/ [–constricted]

It is noteworthy that in spontaneous speech, in stressed and unstressed positions, and in the sentence reading task, in unstressed positions, values for F2 and F3 are quite often reversed, i.e., /ɪ/ takes a higher value than /i/ (see Tables 4.13 and 4.14). This result might indicate a sound change in progress, where /ɪ/ adapts to /i/, in a first step, by neutralizing the degree of constriction (F2, F3), and, in a second step, by neutralizing jaw position (F1). Nevertheless, since at least two speakers uphold the opposition

between /i/ – /ɪ/ in spontaneous speech, assuming a complete neutralization of the pair is not yet justified.

Due to lip protrusion, F2 remains a natural frequency of the front cavity for the pre-palatal rounded vowels. However, only a moderate lip rounding is favourable for the vowel /y/ (Wood 1986) in order to avoid too drastic a lowering of F2. Complementary larynx depression ensures that F2 and F3 come close together. Stevens (1989) even suggests that F2 and F3 come so close together that the spectral peak created by the formant pair is to be regarded as a single-peaked prominence instead of a two-peaked prominence (Stevens 1989: 17). This has in fact been observed for the vowel /y/ in the analysis of the logatomes (see Figure 4.14).

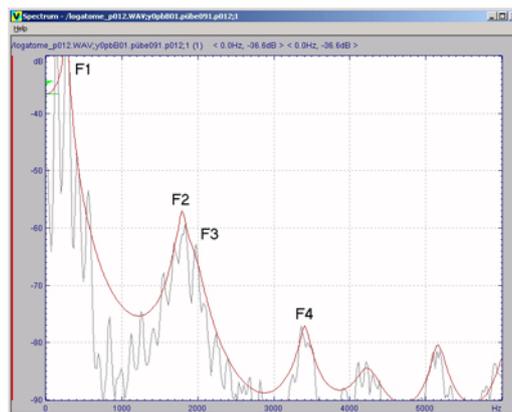


Figure 4.14: Average spectrum of the vowel /y/ taken from the logatome “pübe”, speaker sp012.

Wood (1986) also observed a lower tongue-body for the vowel /y/ as compared to /i/, which diminished the F1 contrast with /i/ (Wood 1986: 396). Pouplier et al (2004) report greater tongue-palate distance only for one out of three subjects. For the others, no relevant distance differences between /i/ and /y/ could be observed.

In the unconstricted /y/, lips are more intensely protruded (Valaczkai 1998: 131), although the ranges overlap. The degree of lip opening, is, as expected, greater for the unconstricted vowel. This very specific lip configuration, together with complementary larynx lowering and difference in tongue-palate distance, leads to a higher F1, a lower F2, and a similar or higher F3 for the unconstricted vowel as compared to the

constricted cognate. Consequently, the spectral peaks of F2 and F3 are clearly kept apart in unconstricted /y/ (see Figure 4.15).

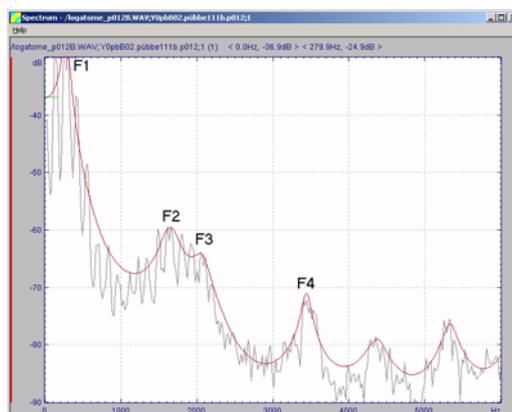


Figure 4.15: Average spectrum of the vowel /y/ taken from the logatome “pübbe”, speaker sp012.

Tongue-palate distance neutralization between the [±constricted] pre-palatal pair /y/ – /ɥ/ occurred only for one subject in Pouplier et al. (2004). However, the small differences in lip configuration, and most probably in complementary larynx lowering, might easily lead to a neutralization as soon as constriction is widened in less formal speech situations or speaking tasks. Unfortunately, in spontaneous speech, all speakers produced either no or too few items for making a statistical comparison reasonable with the exception of speaker sp180 in spontaneous speech in the stressed position. Therefore, in Tables 4.15 – 4.17, under the heading “Ss” and “Sus”, only the mean values are noted (except for sp180, Ss), marked with an asterisk, indicating that no t-test has been performed.

F1	L	L	Rs	Rs	Rus	Rus.	Ss	Ss	Sus	Sus
Sp012	<b>265</b>	<b>290</b>	<b>305</b>	<b>352</b>	<b>315</b>	<b>372</b>	–	–	–	–
Sp180	<b>246</b>	<b>297</b>	<b>304</b>	<b>353</b>	355	328	322	286	318*	358*
Sp082	–	–	<b>378</b>	<b>416</b>	<b>388</b>	<b>451</b>	438*	378*	418*	–
Sp129	–	–	<b>337</b>	<b>407</b>	386	396	383*	406*	374*	–
Sp126	–	–	<b>291</b>	<b>317</b>	312	317	–	316*	337*	317*
Sp127	–	–	<b>252</b>	<b>336</b>	<b>312</b>	<b>339</b>	345*	334*	350*	344*

Table 4.15: Mean F1 values of /y/ and /ɥ/ over all speaking tasks. Within each task, the value to the left represents the vowel /y/, the value to the right the vowel /ɥ/. Statistically

significant differences within each pair ( $p < 0.05$ ) are marked in bold. Where the  $t$ -value is positive, the pair is additionally in italics. The asterisk indicates that no  $t$ -tests have been performed. Legend as in Table 4.12.

F2	L	L	Rs	Rs	Rus	Rus.	Ss	Ss	Sus	Sus
Sp012	<b>1794</b>	<b>1705</b>	1800	1684	<b>1764</b>	<b>1649</b>	–	–	–	–
Sp180	<b>2110</b>	<b>1910</b>	<b>2152</b>	<b>1916</b>	1940	1907	1762	1741	1988*	1821*
Sp082	–	–	<b>1738</b>	<b>1631</b>	<b>1660</b>	<b>1530</b>	1662*	1300*	1500*	–
Sp129	–	–	1837	1792	1811	1767	1731*	1638*	1819*	–
Sp126	–	–	1661	1531	<b>1617</b>	<b>1512</b>	–	1399*	1510*	1618*
Sp127	–	–	1588	1480	1597	1616	1325*	1402*	1309*	1404*

Table 4.16: Mean F2 values of /y/ and /ʏ/ over all speaking tasks. Within each task, the value to the left represents the vowel /y/, the value to the right the vowel /ʏ/. Statistically significant differences within each pair ( $p < 0.05$ ) are marked in bold. Where the  $t$ -value is negative, the pair is additionally in italics. The asterisk indicates that no  $t$ -tests have been performed. Legend as in Table 4.12.

F3	L	L	Rs	Rs	Rus	Rus.	Ss	Ss	Sus	Sus
Sp012	2085	2099	2127	2106	2092	2112	–	–	–	–
Sp180	2502	2570	2783	2657	2589	2693	<b>2439</b>	<b>2583</b>	2538*	2477*
Sp082	–	–	2304	2351	2360	2394	2213*	2400*	2295*	–
Sp129	–	–	<b>2327</b>	<b>2483</b>	2453	2514	2375*	2230*	2466*	–
Sp126	–	–	2134	2139	2117	2158	–	2116*	2220*	2061*
Sp127	–	–	<b>2097</b>	<b>2328</b>	<b>2127</b>	<b>2258</b>	2462*	2391*	2211*	2077*

Table 4.17: Mean F3 values of /y/ and /ʏ/ over all speaking tasks. Within each task, the value to the left represents the vowel /y/, the value to the right the vowel /ʏ/. Statistically significant differences within each pair ( $p < 0.05$ ) are marked in bold. Where the  $t$ -value is positive, the pair is additionally in italics. The asterisk indicates that no  $t$ -tests have been performed. Legend as in Table 4.12.

The opposition between the pair /y/ – /ʏ/ is more endangered than the one between /i/ – /ɪ/, especially because F3 plays a minor role (see Tables 4.15 – 4.17). The relevance of F3 comes into play in less formal or weak positions, when the larynx is raised for the unconstricted vowel, and consequently F3 is raised. This causes statistically significant differences in the spontaneous speech of speaker sp180 and in the unstressed condition of the sentence reading task of speaker sp127.

F2, the correlate of lip protrusion and constriction degree, produces statistically significant differences in the unstressed position in the sentence reading task for at least three speakers. In the same way, F1 is only distinctive for three speakers in unstressed positions.

Since for most speakers, no statistical results are available for spontaneous speech, whether the /y/ – /ʏ/ pair is tending towards neutralization or not cannot ultimately be evaluated. A tentative assumption is that F3 might take over the distinctive role of F2 in less formal or weak positions (see e.g. the results of sp127 in “Rus” and of sp180 in “Ss”). Such a change across speaking tasks, situations and prosodic positions also implies a change in the articulatory settings, and consequently in the articulatory features. In other words, the feature [±constricted] would be replaced by a feature larynx height. This change results from the very subtle interplay of degree of lip protrusion with accompanying degree of larynx lowering, degree of lip aperture, and degree of constriction. Since, however, tongue-palate distance is still relevant in the unstressed position of the sentence reading task for at least three speakers, the feature [±constricted] is maintained.

#### **4.4.2. The mid-palatal vowels**

If the tongue forms a constriction in the mid-palatal region, F2 is a natural frequency of the front cavity, and, therefore, relatively high, and F3 is a natural frequency of the back cavity and, consequently, substantially lower than for /i/ (see Figure 4.16).

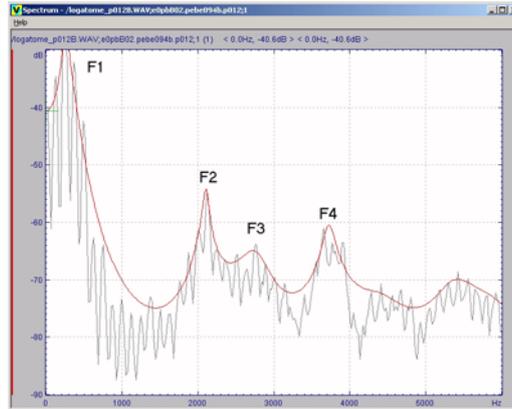


Figure 4.16: Average spectrum of the vowel /e/ taken from the logatome “pebe”, speaker sp012.

Since F2 is relatively high and F3 is maximally low, these two frequencies can form a center of gravity. For the unconstricted counterpart in the mid-palatal location / $\epsilon$ /, the degree of constriction is even more widened than for unconstricted /i/. One can therefore assume a great acoustic coupling which yields in a high F1 (together with a high degree of lip aperture) and a substantial lowering of both F2 and F3 (see Figure 4.17).

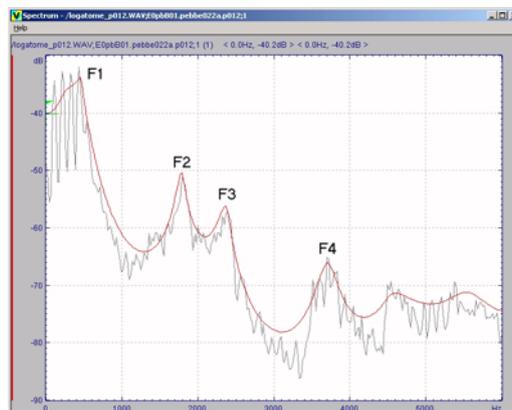


Figure 4.17: Average spectrum of the vowel / $\epsilon$ / taken from the logatome “pebbe”, speaker sp012.

The vowel pair /e/ and / $\epsilon$ / is primarily discerned by constriction degree, which is substantially greater for the unconstricted vowel / $\epsilon$ /. Wood (1982) reports approximately the same degree of lip aperture for both vowels, whereas in Valaczkai (1998), the degree of lip aperture is greater for the unconstricted vowel / $\epsilon$ /. These differences in the articulatory settings between /e/ and / $\epsilon$ / lead to a higher F1, a lower F2, and a lower

F3 for the unconstricted vowel /e/ as compared to the constricted cognate /ɛ/ (see Tables 4.18 – 4.20).

F1	L	L	Rs	Rs	Rus	Rus.	Ss	Ss	Sus	Sus
Sp012	<b>316</b>	<b>434</b>	<b>331</b>	<b>444</b>	376	387	<b>373</b>	<b>436</b>	<b>364</b>	<b>431</b>
Sp180	<b>336</b>	<b>434</b>	<b>347</b>	<b>422</b>	369	376	<b>353</b>	<b>424</b>	396	386
Sp082	–	–	<b>392</b>	<b>494</b>	<b>415</b>	<b>432</b>	<b>393</b>	<b>491</b>	<b>410</b>	<b>446</b>
Sp129	–	–	<b>366</b>	<b>461</b>	<b>429</b>	<b>438</b>	<b>393</b>	<b>444</b>	<b>414</b>	<b>441</b>
Sp126	–	–	<b>323</b>	<b>404</b>	<b>346</b>	<b>359</b>	<b>346</b>	<b>403</b>	<b>339</b>	<b>362</b>
Sp127	–	–	<b>330</b>	<b>417</b>	<b>351</b>	<b>366</b>	<b>348</b>	<b>417</b>	352	364

Table 4.18: Mean F1 values of /e/ and /ɛ/ over all speaking tasks. Within each task, the value to the left represents the vowel /e/, the value to the right the vowel /ɛ/. Statistically significant differences within each pair ( $p < 0.05$ ) are marked in bold. Where the  $t$ -value is positive, the pair is additionally in italics. Legend as in Table 4.12.

F2	L	L	Rs	Rs	Rus	Rus.	Ss	Ss	Sus	Sus
Sp012	<b>2097</b>	<b>1793</b>	<b>2080</b>	<b>1807</b>	<b>1747</b>	<b>1663</b>	<b>1936</b>	<b>1704</b>	1630	1551
Sp180	<b>2562</b>	<b>2244</b>	<b>2421</b>	<b>2229</b>	<b>2135</b>	<b>2024</b>	<b>2441</b>	<b>2205</b>	2037	1979
Sp082	–	–	<b>2410</b>	<b>2056</b>	<b>1891</b>	<b>1844</b>	<b>2184</b>	<b>2003</b>	<b>1873</b>	<b>1764</b>
Sp129	–	–	<b>2461</b>	<b>2209</b>	<b>1997</b>	<b>1873</b>	<b>2293</b>	<b>2058</b>	<b>2044</b>	<b>1789</b>
Sp126	–	–	<b>2095</b>	<b>1797</b>	<b>1713</b>	<b>1619</b>	<b>1911</b>	<b>1704</b>	1597	1549
Sp127	–	–	<b>2086</b>	<b>1911</b>	<b>1777</b>	<b>1682</b>	<b>1924</b>	<b>1767</b>	<b>1744</b>	<b>1582</b>

Table 4.19: Mean F2 values of /e/ and /ɛ/ over all speaking tasks. Within each task, the value to the left represents the vowel /e/, the value to the right the vowel /ɛ/. Statistically significant differences within each pair ( $p < 0.05$ ) are marked in bold. Where the  $t$ -value is negative, the pair is additionally in italics. Legend as in Table 4.12.

F3	L	L	Rs	Rs	Rus	Rus.	Ss	Ss	Sus	Sus
Sp012	<b>2817</b>	<b>2351</b>	<b>2618</b>	<b>2303</b>	<b>2267</b>	<b>2225</b>	<b>2478</b>	<b>2157</b>	2198	2157
Sp180	<b>3182</b>	<b>2818</b>	<b>3071</b>	<b>2852</b>	2774	2795	<b>2993</b>	<b>2769</b>	<b>2665</b>	<b>2713</b>
Sp082	–	–	<b>2939</b>	<b>2516</b>	<b>2432</b>	<b>2466</b>	2589	2618	2427	2459
Sp129	–	–	<b>3026</b>	<b>2716</b>	<b>2615</b>	<b>2581</b>	<b>2710</b>	<b>2569</b>	<b>2568</b>	<b>2496</b>
Sp126	–	–	<b>2557</b>	<b>2327</b>	<b>2331</b>	<b>2262</b>	<b>2431</b>	<b>2315</b>	2222	<b>2323</b>
Sp127	–	–	<b>2625</b>	<b>2476</b>	<b>2386</b>	<b>2452</b>	<b>2511</b>	<b>2457</b>	<b>2365</b>	<b>2452</b>

Table 4.20: Mean F3 values of /e/ and /ɛ/ over all speaking tasks. Within each task, the value to the left represents the vowel /e/, the value to the right the vowel /ɛ/. Statistically significant differences within each pair ( $p < 0.05$ ) are marked in bold. Where the  $t$ -value is negative, the pair is additionally in italics. Legend see Table 4.12.

As becomes apparent from Tables 4.18 – 4.20, the two vowels /e/ and /ɛ/ are clearly kept apart. In the logatome reading task, both speakers discern the pair by all three formants, and in the sentence reading task in the stressed position, all speakers but one discern the pair by all three formants. Figure 4.18 summarizes the results of the statistical analysis:

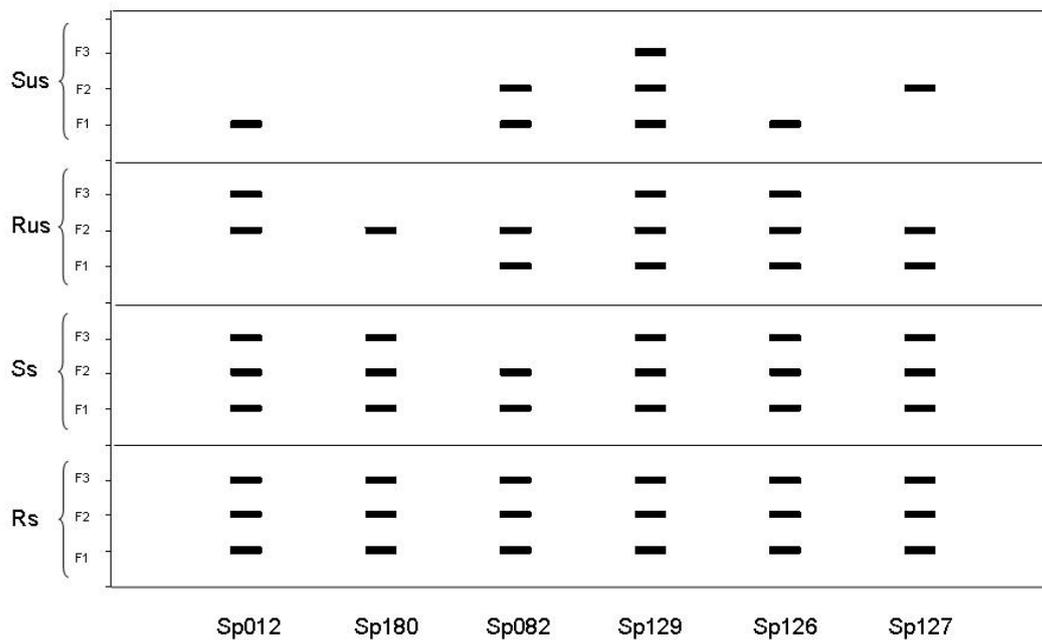


Figure 4.18: Results of the one-tailed t-tests for the vowel pair /e/ – /ɛ/. For each formant and each speaking task, statistically significant differences ( $p < 0.05$ ) are indicated by crossbeams. Where no differences occur for a given formant, the space is left blank. Legend as in Table 4.12.

It can be read from Figure 4.18 that an opposition is maintained for at least one formant in each speaking task (except for speaker sp180 who neutralizes the opposition in the unstressed position in spontaneous speech). However, even more important is the result that in unstressed positions, speakers maintain the opposition in different ways. Apart from the fact that reveals some speaker-specific information, this result also points out that speakers make use of different articulatory configurations and consequently of

different features to maintain an opposition. It should also be emphasized that a neutralization of F1 (speaker sp180 and speaker sp127) leads to a lowering of the formant of the unconstricted vowel /ɛ/, and not to a rising of the F1 of the constricted vowel /e/. In other words, the unconstricted vowels assimilates to the constricted one<sup>78</sup> in neutralization, and not the other way round.

However, given the non-linear relationship between articulation and acoustics, different articulatory configurations might lead to the same acoustic output. In Standard Austrian German, F2 of /i/ and /e/ might be affected. It has already been argued (Chapter 3 and 4.4.1) that the pre-palatal location of /i/ leads to a shift in cavity affiliation and, consequently, to a drastic rise in F3. It has also been argued, and it can be read from Table 4.14, that F3 drops drastically as soon as constriction degree is widened or constriction length is shortened. F3 of /e/ has, due to mid-palatal constriction location and concomitant back cavity affiliation, a lower value than in /i/, forming a center of gravity with F2 (see Figure 4.15). Since F3 of /i/ is relatively unstable, and since F2 of /i/ either has lower values than /e/ or identical values to /e/, the opposition of /i/ and /e/ might not be upheld over all speaking tasks and prosodic positions. Therefore, it is necessary to test whether an acoustic overlap between /i/ and /e/ takes place in any of the analyzed speaking tasks or prosodic positions. Tables 4.21 – 4.23 present the results.

F1	L	L	Rs	Rs	Rus	Rus.	Ss	Ss	Sus	Sus
Sp012	<b>269</b>	<b>316</b>	<b>284</b>	<b>331</b>	<b>317</b>	<b>376</b>	<b>330</b>	<b>373</b>	<b>337</b>	<b>364</b>
Sp180	<b>241</b>	<b>336</b>	<b>301</b>	<b>347</b>	<b>320</b>	<b>369</b>	<b>306</b>	<b>353</b>	<b>341</b>	<b>396</b>
Sp082	–	–	379	392	<b>398</b>	<b>415</b>	<b>363</b>	<b>393</b>	<b>383</b>	<b>410</b>
Sp129	–	–	<b>355</b>	<b>366</b>	<b>382</b>	<b>429</b>	<b>367</b>	<b>393</b>	<b>373</b>	<b>414</b>
Sp126	–	–	<b>282</b>	<b>323</b>	<b>308</b>	<b>346</b>	<b>302</b>	<b>346</b>	<b>325</b>	<b>339</b>
Sp127	–	–	<b>284</b>	<b>330</b>	<b>307</b>	<b>351</b>	<b>311</b>	<b>348</b>	<b>320</b>	<b>352</b>

Table 4.21: Mean F1 values of /i/ and /e/ over all speaking tasks. Within each task, the value to the left represents the vowel /i/, the value to the right the vowel /e/. Statistically

<sup>78</sup> A fact that challenges the undershoot hypothesis, see also Chapter 5.

significant differences within each pair ( $p < 0.05$ ) are marked in bold. Where the  $t$ -value is positive, the pair is additionally in italics. Legend as in Table 4.12.

F2	L	L	Rs	Rs	Rus	Rus.	Ss	Ss	Sus	Sus
Sp012	2078	2097	<b>2054</b>	<b>2080</b>	<b>1918</b>	<b>1747</b>	<b>2079</b>	<b>1936</b>	<b>1758</b>	<b>1630</b>
Sp180	2533	2562	2489	2421	<b>2256</b>	<b>2135</b>	2350	2441	<b>2153</b>	<b>2037</b>
Sp082	–	–	<b>2334</b>	<b>2410</b>	<b>2059</b>	<b>1891</b>	2179	2184	1907	1873
Sp129	–	–	<b>2369</b>	<b>2461</b>	<b>2183</b>	<b>1997</b>	2379	2293	<b>2116</b>	<b>2044</b>
Sp126	–	–	2094	2095	<b>1813</b>	<b>1713</b>	1985	1911	<b>1744</b>	<b>1597</b>
Sp127	–	–	2110	2086	<b>1936</b>	<b>1777</b>	<b>2043</b>	<b>1924</b>	1705	1744

Table 4.22: Mean F2 values of /i/ and /e/ over all speaking tasks. Within each task, the value to the left represents the vowel /i/, the value to the right the vowel /e/. Statistically significant differences within each pair ( $p < 0.05$ ) are marked in bold. Two-tailed  $t$ -test have been performed in this case, since either direction is possible for F2. Legend as in Table 4.12.

F3	L	L	Rs	Rs	Rus	Rus.	Ss	Ss	Sus	Sus
Sp012	<b>3230</b>	<b>2817</b>	<b>2828</b>	<b>2618</b>	<b>2508</b>	<b>2267</b>	<b>2700</b>	<b>2478</b>	<b>2305</b>	<b>2198</b>
Sp180	<b>3413</b>	<b>3182</b>	<b>3179</b>	<b>3071</b>	<b>2916</b>	<b>2774</b>	<b>3042</b>	<b>2993</b>	<b>2781</b>	<b>2665</b>
Sp082	–	–	2893	2939	<b>2509</b>	<b>2432</b>	<b>2733</b>	<b>2589</b>	2475	2427
Sp129	–	–	3063	3026	<b>2782</b>	<b>2615</b>	<b>2993</b>	<b>2710</b>	<b>2656</b>	<b>2568</b>
Sp126	–	–	<b>2697</b>	<b>2557</b>	2351	2331	<b>2561</b>	<b>2431</b>	2270	2222
Sp127	–	–	2674	2625	<b>2439</b>	<b>2386</b>	2516	2511	2403	2365

Table 4.23: Mean F3 values of /i/ and /e/ over all speaking tasks. Within each task, the value to the left represents the vowel /i/, the value to the right the vowel /e/. Statistically significant differences within each pair ( $p < 0.05$ ) are marked in bold. Where the  $t$ -value is negative, the pair is additionally in italics. Legend as in Table 4.12.

As becomes evident from Table 4.21, there is absolutely no question that /i/ and /e/ are discerned by F1, i.e. degree of lip aperture, over all speaking tasks and prosodic positions. The opposition concerning F3, the most relevant parameter for constriction location, is only upheld by speaker sp012 and speaker sp180 over all tasks and prosodic positions. Speaker sp126 still maintains the opposition in stressed positions, but shifts the distinction to F2. The same holds for the stressed position for speakers sp082 and sp129 in the sentence reading task. It has to be noted that for these two speakers F2 takes higher values for /e/ than for /i/, a fact that clearly points to the pre-palatal location of /i/. Speaker sp012 also shows higher F2 values for /e/ than for /i/, but he discerns F3

as well. The only speaker who does not fit into this pattern is speaker sp127, the youngest speaker of all, who, in the stressed positions in the sentence reading task, neither upholds an opposition with respect to F2, nor with respect to F3. The same holds for unstressed positions for this speaker and speaker sp082 in spontaneous speech. It cannot be decided at this point of the investigation, whether these deviations point to a sound change which changes the feature [±pre-palatal] to [±open], leaving the discriminatory ability to F1 only. Anyhow, five out of six speakers clearly discern two constriction locations in the front region.

According to perception tests (Linder 1976, Sendlmeier 1981), /e/ tends to be mixed up with /ɪ/. These results lead Sendlmeier (1985) to see /e/ rather as the long partner of /ɪ/ than of /ɛ/. Consequently, it has to be tested whether these two vowels are held apart acoustically in Standard Austrian German.

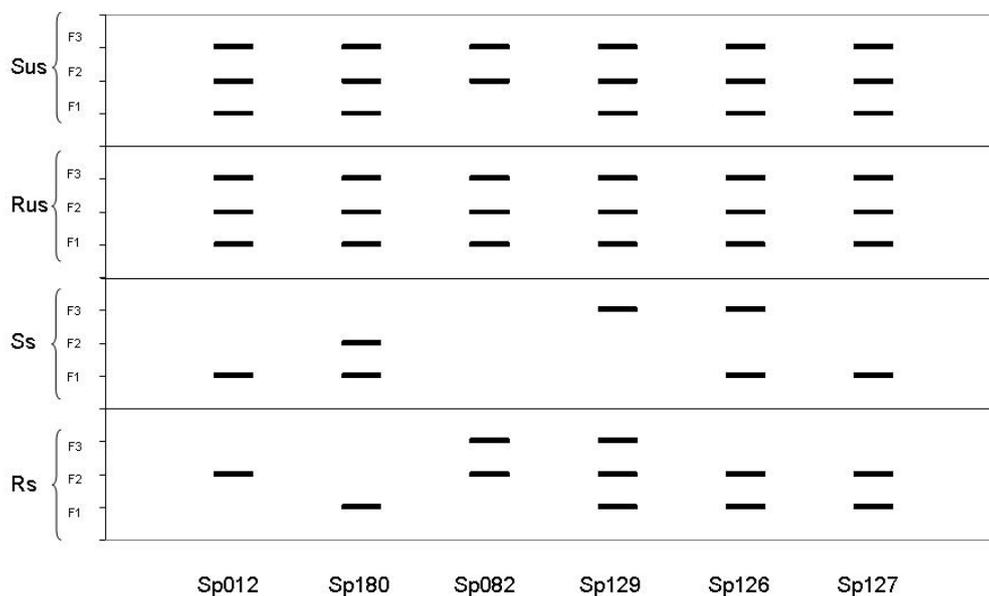


Figure 4.19: Results of the one-tailed t-tests for the vowel pair /e/ - /ɪ/. For each formant and each speaking task, statistically significant differences ( $p < 0.05$ ) are indicated by crossbeams. Where no differences occur for a given formant, the space is left blank. Legend as in Table 4.12.

It can be read from Figure 4.19 that, with the exception of speaker sp082, all speakers differentiate /e/ – /ɪ/ at least with one formant in spontaneous speech in the stressed position. These results hold also for the logatome reading task. Therefore, in Standard Austrian German, these two vowels are clearly kept apart. It is interesting, however, that for unstressed positions in the sentence reading task, all speakers discriminate the two vowels by all three formants, and for unstressed positions in spontaneous speech, five speakers discriminate the two vowels by means of all three formants, and one speaker (sp082) by means of F2 and F3. This means that, contrary to expectations and contrary to the results presented so far, the discriminatory power is stronger in unstressed positions. Although no explanation can be presented for this behaviour, it clearly shows, at least, that /e/ is by no means the cognate of /ɪ/.

For the mid-palatal vowel /ø/, moderate lip protrusion with compensatory larynx lowering leads to an approximation of F2 and F3. Contrary to /y/, the two peaks could, however, be dissolved in most cases (see Figure 4.20).

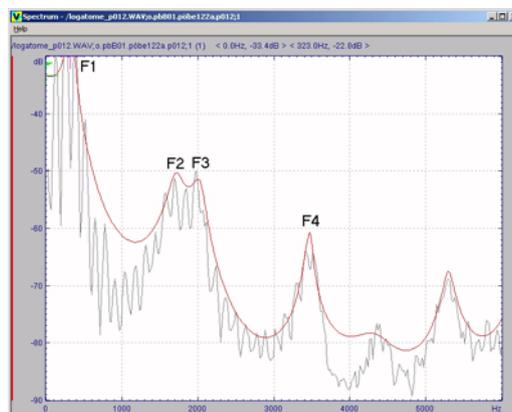


Figure 4.20: Average spectrum of the vowel /ø/ taken from the logatome “pöbe”, speaker sp012.

Following the articulatory description given in Valaczkai (1998), the unconstricted /œ/ is – as compared to its constricted cognate – produced with a higher degree of lip opening (although some overlap is reported), an approximately equal degree of lip protrusion, and a wider constriction degree<sup>79</sup>, leading to a higher coupling of the front

<sup>79</sup> In Pouplier et al. (2004), no tongue-palate distances for the constricted vowel /ø/ are given.

and back cavity. Since lip protrusion is about the same for both vowels, the larynx should be lowered by roughly the same degree as well, or, to a lesser degree by the unconstricted vowel. These articulatory configurations lead to a higher F1, a lower F2, and an equal or higher F3 for unconstricted /œ/ as compared to its constricted cognate /ø/ (see Figure 4.21)

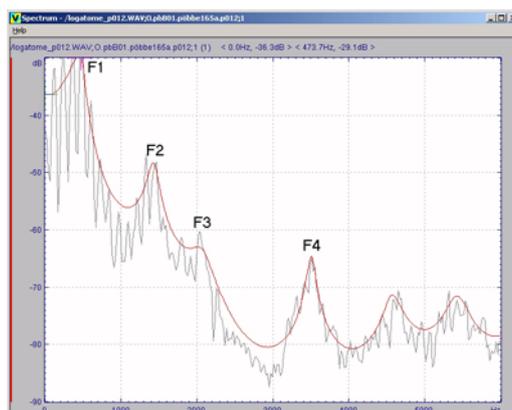


Figure 4.21: Average spectrum of the vowel /œ/ taken from the logatome “pöbbe”, speaker sp012.

Tables 4.24 – 4.26 give the statistical results of the pair-wise comparisons. Unfortunately, especially in spontaneous speech, not enough items were produced to make a statistical comparison meaningful. Therefore, in the same way as for the vowel pair /y/ – /ʏ/, mean values, if available, are presented and supplied with an asterisk in order to indicate that no statistical analysis could be performed for this pair.

F1	L	L	Rs	Rs	Rus	Rus.	Ss	Ss	Sus	Sus
Sp012	<b>324</b>	<b>436</b>	<b>347</b>	<b>439</b>	354*	493*	353*	441*	–	–
Sp180	<b>351</b>	<b>452</b>	<b>394</b>	<b>430</b>	–	364*	387*	–	–	–
Sp082	–	–	<b>423</b>	<b>502</b>	–	–	424*	–	–	–
Sp129	–	–	<b>391</b>	<b>486</b>	–	455*	421*	443*	450*	432*
Sp126	–	–	<b>342</b>	<b>393</b>	320*	–	–	406*	–	–
Sp127	–	–	<b>342</b>	<b>436</b>	354*	476*	326*	–	–	–

Table 4.24: Mean F1 values of /ø/ and /œ/ over all speaking tasks. Within each task, the value to the left represents the vowel /ø/, the value to the right the vowel /œ/. Statistically significant differences within each pair ( $p < 0.05$ ) are marked in bold. Where the t-value is positive, the pair is additionally in italics. The asterisk indicates that no t-tests have been performed. Legend as in Table 4.12.

F2	L	L	Rs	Rs	Rus	Rus.	Ss	Ss	Sus	Sus
Sp012	<b>1701</b>	<b>1482</b>	<b>1669</b>	<b>1560</b>	1464*	1616*	1567*	1648*	–	–
Sp180	<b>1896</b>	<b>1811</b>	1949	1932	–	2013*	1877*	–	–	–
Sp082	–	–	1651	1628	–	–	1627*	–	–	–
Sp129	–	–	1749	1669	–	1868*	1562*	1633*	1724*	1801*
Sp126	–	–	1506	1478	1456*	–	–	1368*	–	–
Sp127	–	–	1508	1550	1488*	1450*	1552*	–	–	–

Table 4.25: Mean F2 values of /ø/ and /œ/ over all speaking tasks. Within each task, the value to the left represents the vowel /ø/, the value to the right the vowel /œ/. Statistically significant differences within each pair ( $p < 0.05$ ) are marked in bold. Where the t-value is negative, the pair is additionally in italics. The asterisk indicates that no t-tests have been performed. Legend as in Table 4.12.

F3	L	L	Rs	Rs	Rus	Rus.	Ss	Ss	Sus	Sus
Sp012	2021	2043	<b>2010</b>	<b>2067</b>	2054*	2119*	2038*	2012*	–	–
Sp180	2549	2563	2765	2714	–	2970*	2577*	–	–	–
Sp082	–	–	<b>2313</b>	<b>2354</b>	–	–	2372*	–	–	–
Sp129	–	–	2433	2463	–	2354*	2442*	2442*	2636*	2362*
Sp126	–	–	<b>2052</b>	<b>2140</b>	2041*	–	–	2012*	–	–
Sp127	–	–	<b>2131</b>	<b>2244</b>	2283*	2140*	2224*	–	–	–

Table 4.26: Mean F3 values of /ø/ and /œ/ over all speaking tasks. Within each task, the value to the left represents the vowel /ø/, the value to the right the vowel /œ/. Statistically significant differences within each pair ( $p < 0.05$ ) are marked in bold. Where the t-value is positive, the pair is additionally in italics. The asterisk indicates that no t-tests have been performed. Legend as in Table 4.12.

All six speakers discern /ø/ and /œ/ by F1. In other words, the degree of lip opening plays a decisive role in the discriminability of the vowel pair. As expected, F2 is only discerned by one speaker in stressed positions in the sentence reading task, whereas F3 is still discerned by four speakers, rendering, as expected, a higher value for the unconstricted vowel /œ/. Since there are no data available for spontaneous speech, it cannot be evaluated whether these oppositions are maintained over all speaking tasks and prosodic positions, or whether a tendency towards neutralization, due to the only moderate lip protrusion and the – probably moderate – accompanying larynx lowering for the constricted vowel, is on the way.

### 4.4.3. The velar vowels

Ideally, the /u/ – vowels are discerned by constriction degree and lip protrusion (Wood 1975b). However, for German, Valaczkai (1998) found no differences in the degree of lip protrusion between /u/ and /ʊ/, but reports a slightly higher degree of lip opening and a higher degree of constriction for the unconstricted vowel /ʊ/. As regards tongue-palate distance, Pouplier et al. (2004) found a sufficient difference (> 1 mm) only for one subject out of three. A constriction in the velar region leads to a lowering of F2 and a rising of F3 (Fant 2004: 43). Lip protrusion would lower all three formants. Therefore, an articulatory configuration which bases the opposition on the degree of constriction would lead to a higher F1, a higher F2, and a lower F3 for the unconstricted vowel /ʊ/, whereas a configuration which bases the opposition on lip protrusion would lead to a higher F1, F2, and F3 for the unconstricted vowel /ʊ/. The two spectra presented in Figures 4.22 and 4.23 rather point to a difference in lip protrusion, since F2 and F3 are both higher for /ʊ/ than for /u/.

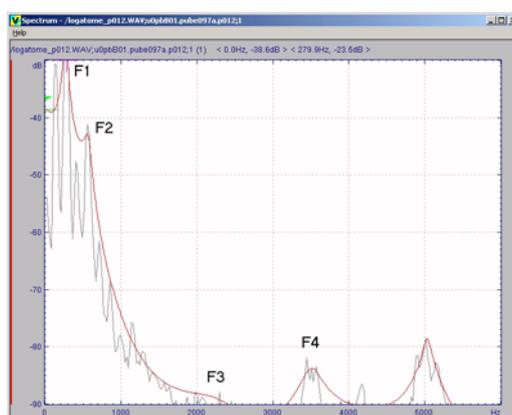


Figure 4.22: Average spectrum of the vowel /u/ taken from the logatome “pube”, speaker sp012.

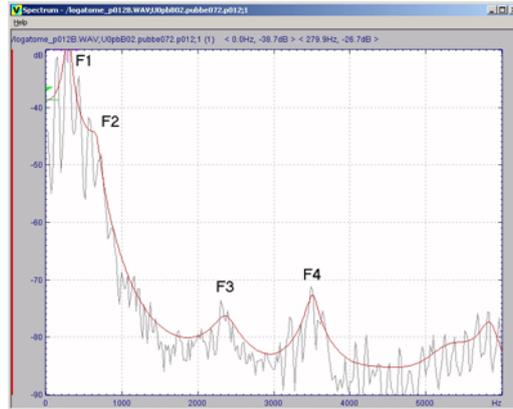


Figure 4.23: Average spectrum of the vowel /u/ taken from the logatome “pubbe”, speaker sp012.

However, statistical results offer no clear picture, and it might well be the case that articulatory configurations are speaker-dependent (see Tables 4.27 – 4.29).

F1	L	L	Rs	Rs	Rus	Rus.	Ss	Ss	Sus	Sus
Sp012	<b>279</b>	<b>357</b>	<b>313</b>	<b>345</b>	334	349	321*	402*	351*	364*
Sp180	<b>262</b>	<b>311</b>	<b>315</b>	<b>346</b>	398	371	347	312	367	371
Sp082	–	–	<b>372</b>	<b>429</b>	400	402	384*	359*	409	441
Sp129	–	–	<b>343</b>	<b>378</b>	407	405	366	380	394	413
Sp126	–	–	<b>308</b>	<b>330</b>	329	323	337	336	340	333
Sp127	–	–	<b>301</b>	<b>351</b>	<b>329</b>	<b>350</b>	329	333	341	336

Table 4.27: Mean F1 values of /u/ and /u/ over all speaking tasks. Within each task, the value to the left represents the vowel /u/, the value to the right the vowel /u/. Statistically significant differences within each pair ( $p < 0.05$ ) are marked in bold. Where the t-value is positive, the pair is additionally in italics. The asterisk indicates that no t-tests have been performed. Legend as in Table 4.12.

F2	L	L	Rs	Rs	Rus	Rus.	Ss	Ss	Sus	Sus
Sp012	<b>626</b>	<b>826</b>	<b>654</b>	<b>723</b>	1065	1227	1020*	1124*	1137*	1281*
Sp180	<b>887</b>	<b>969</b>	953	1060	<b>1437</b>	<b>1247</b>	1162	1096	<b>1393</b>	<b>1199</b>
Sp082	–	–	<b>867</b>	<b>1046</b>	<b>1314</b>	<b>1293</b>	975*	1119*	<b>1507</b>	<b>998</b>
Sp129	–	–	747	752	<b>1098</b>	<b>880</b>	836	837	1202	1038
Sp126	–	–	<b>768</b>	<b>916</b>	1140	1183	<b>930</b>	<b>779</b>	<b>1179</b>	<b>1066</b>
Sp127	–	–	814	869	1240	1221	922	814	1107	1112

Table 4.28: Mean F2 values of /u/ and /u/ over all speaking tasks. Within each task, the value to the left represents the vowel /u/, the value to the right the vowel /u/. Statistically significant differences within each pair ( $p < 0.05$ ) are marked in bold. Where a t-

value is positive, the pair is additionally in italics. The asterisk indicates that no t-tests have been performed. Legend as in Table 4.12.

F3	L	L	Rs	Rs	Rus	Rus.	Ss	Ss	Sus	Sus
Sp012	2411	2419	2128	2080	2099	2077	2381*	2235*	2249*	2300*
Sp180	2533	2454	2612	2460	2510	2541	2603	2654	2630	2693
Sp082	–	–	2383	2455	2443	2479	2491*	2712*	2440	2520
Sp129	–	–	<b>2659</b>	<b>2504</b>	<b>2586</b>	<b>2460</b>	2361	2319	<b>2523</b>	<b>2407</b>
Sp126	–	–	2221	2214	2203	2258	<b>2260</b>	<b>2149</b>	2299	2299
Sp127	–	–	2345	2330	2393	2291	2428	2371	2440	2324

Table 4.29: Mean F3 values of /u/ and /ʊ/ over all speaking tasks. Within each task, the value to the left represents the vowel /u/, the value to the right the vowel /ʊ/. Statistically significant differences within each pair ( $p < 0.05$ ) are marked in bold. Where the t-value is negative, the pair is additionally in italics. The asterisk indicates that no t-tests have been performed. Legend as in Table 4.12.

It can be seen that an opposition in a stressed position is only maintained in the most formal speaking tasks of reading logatomes and sentences, predominately for F1. F2 produces a very inconsistent picture, often with reversed values (higher values for the [+constricted] vowel), especially in unstressed positions. Only one speaker (sp129) maintains an opposition for F3 over all speaking tasks (except in stressed positions in spontaneous speech). Given the reversed values for F2 of this speaker in unstressed positions in the sentence reading task, the corresponding F3 values rather point to a reduction of the degree of lip protrusion (both F2 and F3 are higher for /u/ than for /ʊ/) than to a difference in constriction degree. The same holds for speaker sp126, for the stressed position in spontaneous speech. In all other cases, no statistically significant differences were obtained for F3. Together with the results obtained for F2 and F1, it has to be concluded that neutralization of the velar vowels is even more advanced than for the pre-palatal vowels.

This assumption is corroborated by the fact that even in the logatome reading task, speaker sp012 clearly realizes some /ʊ/ vowels as /u/. The items affected are: “kugge”,

“gugge”, gukke” (2x), “puppe”, and “kukke”<sup>80</sup>. T-tests render no differences with the vowel /u/ for F1 and F2, and a slight difference for F3, the value being lower for the phonologically [–constricted] vowel, pointing to a difference in constriction degree.

#### 4.4.4. The upper pharyngeal vowels

A constriction in the upper pharyngeal region leads to a low F1, low F2, and a high F3. Accompanied by lip protrusion, a low F2 value is maintained over long stretches of constriction locations (quantal region, Stevens 1972, 1989); consequently, there is hardly any difference in F2 between /u/ and /o/. Acoustically, the difference between /u/ and /o/ is carried out by F1 and F3, which are both higher for /o/ (see Figure 4.24 as compared to Figure 4.22).

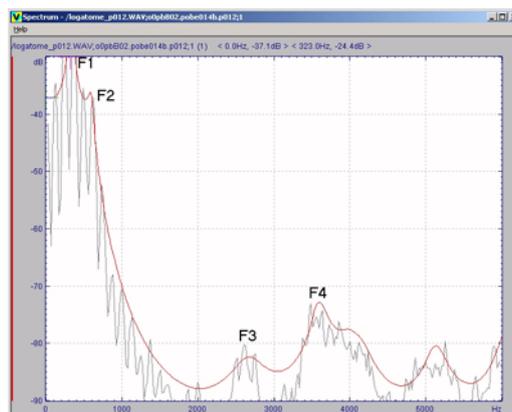


Figure 4.24: Average spectrum of the vowel /o/ taken from the logatome “pobe”, speaker sp012.

Both constriction degree and degree of lip opening are higher for the unconstricted cognate /ɔ/. Degree of lip protrusion, however, does not differ substantially (Valaczkai 1998). Pouplier et al. (2004) report higher tongue-palate distance for unconstricted /ɔ/. Consequently, F1 and F2 are higher for /ɔ/, and F3 is lower than for /o/ (Figure 4.25).

<sup>80</sup> Interestingly, Pouplier et al. (2004), who observed neutralization of tongue-palate distance for the /u/-vowels, analyzed their vowels in velar context.

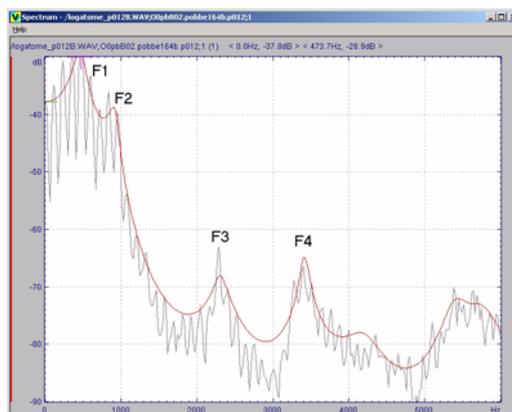


Figure 4.25: Average spectrum of the vowel /ɔ/ taken from the logatome “pobbe”, speaker sp012.

Tables 4.30 – 4.32 present the statistical results for the pair-wise comparison of /o/ and /ɔ/.

F1	L	L	Rs	Rs	Rus	Rus.	Ss	Ss	Sus	Sus
Sp012	<b>342</b>	<b>486</b>	<b>368</b>	<b>512</b>	<b>387</b>	<b>475</b>	424	458	383	459
Sp180	<b>358</b>	<b>490</b>	<b>375</b>	<b>495</b>	420	404	<b>386</b>	<b>443</b>	399	390
Sp082	–	–	<b>409</b>	<b>534</b>	<b>464</b>	<b>499</b>	<b>430</b>	<b>515</b>	423	457
Sp129	–	–	<b>377</b>	<b>478</b>	<b>432</b>	<b>461</b>	<b>395</b>	<b>432</b>	<b>421</b>	<b>454</b>
Sp126	–	–	<b>348</b>	<b>425</b>	<b>355</b>	<b>385</b>	<b>380</b>	<b>416</b>	<b>362</b>	<b>388</b>
Sp127	–	–	<b>351</b>	<b>469</b>	<b>364</b>	<b>414</b>	<b>366</b>	<b>410</b>	<b>364</b>	<b>402</b>

Table 4.30: Mean F1 values of /o/ and /ɔ/ over all speaking tasks. Within each task, the value to the left represents the vowel /o/, the value to the right the vowel /ɔ/. Statistically significant differences within each pair ( $p < 0.05$ ) are marked in bold. Where the t-value is positive, the pair is additionally in italics. Legend as in Table 4.12.

F2	L	L	Rs	Rs	Rus	Rus.	Ss	Ss	Sus	Sus
Sp012	<b>620</b>	<b>884</b>	<b>646</b>	<b>846</b>	<b>829</b>	<b>1012</b>	<b>722</b>	<b>997</b>	1116	1016
Sp180	<b>842</b>	<b>1049</b>	<b>854</b>	<b>1136</b>	1253	1290	965	1128	<b>1339</b>	<b>1226</b>
Sp082	–	–	<b>828</b>	<b>1083</b>	<b>1224</b>	<b>1289</b>	<b>867</b>	<b>964</b>	1157	1229
Sp129	–	–	<b>671</b>	<b>870</b>	<b>906</b>	<b>999</b>	773	792	1194	1093
Sp126	–	–	<b>736</b>	<b>928</b>	<b>962</b>	<b>1126</b>	<b>727</b>	<b>968</b>	1034	1024
Sp127	–	–	<b>730</b>	<b>844</b>	<b>970</b>	<b>1133</b>	819	907	1086	1066

Table 4.31: Mean F2 values of /o/ and /ɔ/ over all speaking tasks. Within each task, the value to the left represents the vowel /o/, the value to the right the vowel /ɔ/. Statistically significant differences within each pair ( $p < 0.05$ ) are marked in bold. Where the t-value is positive, the pair is additionally in italics. Legend as in Table 4.12.

F3	L	L	Rs	Rs	Rus	Rus.	Ss	Ss	Sus	Sus
Sp012	<b>2525</b>	<b>2125</b>	<b>2107</b>	<b>1988</b>	<b>2124</b>	<b>2042</b>	2144	2094	2176	2133
Sp180	2649	2748	2675	2575	2651	2615	<b>2761</b>	<b>2588</b>	2642	2592
Sp082	–	–	2438	2479	2483	2448	2471	2431	2515	2538
Sp129	–	–	<b>2674</b>	<b>2361</b>	<b>2550</b>	<b>2451</b>	2459	2370	<b>2503</b>	<b>2270</b>
Sp126	–	–	<b>2244</b>	<b>1938</b>	2147	2114	<b>2228</b>	<b>2110</b>	<b>2244</b>	<b>2080</b>
Sp127	–	–	2454	2420	2338	2277	2481	2330	2496	2374

Table 4.32: Mean F3 values of /o/ and /ɔ/ over all speaking tasks. Within each task, the value to the left represents the vowel /o/, the value to the right the vowel /ɔ/. Statistically significant differences within each pair ( $p < 0.05$ ) are marked in bold. Where the t-value is negative, the pair is additionally in italics. Legend as in Table 4.12.

In reading logatomes, both speakers discern the pair for F1 and F2. F3 is only used by speaker sp012. The same picture emerges for the stressed position in the sentence reading task; all speakers discern the pair for F1 and F2, but only three use F3 as well. Figure 4.26 summarizes the results of the statistical analysis:

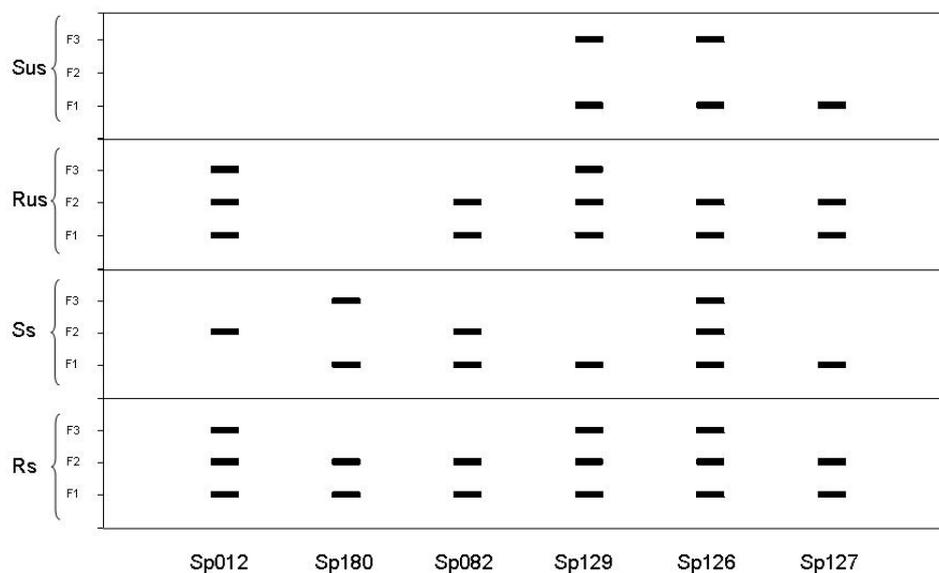


Figure 4.26: Results of the one-tailed t-tests for the vowel pair /o/ – /ɔ/. For each formant and each speaking task, statistically significant differences ( $p < 0.05$ ) are indicated by crossbeams. Where no differences occur for a given formant, the space is left blank. Legend as in Table 4.12.

Speaker-specific maintenance of the opposition in the stressed position already occurs in spontaneous speech (see Figure 4.26). In unstressed positions, three speakers neutralize the opposition in spontaneous speech, and one speaker in the sentence reading task. Figure 4.26 also reveals that F1 plays an important role in discriminating the two vowels, where the difference with respect to F1 is the last to be discarded. In spontaneous speech, the discriminatory ability of F2 is less pronounced. F3, however, plays an additional role in discriminating the two vowels. Two speakers maintain a difference in unstressed positions in spontaneous speech, and two in stressed positions in spontaneous speech. F3 values are lower for the [–constricted] vowel throughout, pointing to a higher constriction degree for the vowel /ɔ/. It has to be noted, however, that the most stable discriminatory parameter is F1, pointing to a higher degree of lip opening for /ɔ/, whereas constriction degree starts to get neutralized in spontaneous speech, thus corroborating the results of Wood (1975b).

#### **4.4.5. The lower pharyngeal vowels**

The lower pharyngeal vowels are traditionally described as front /a/ or back /ɑ/. The description of front or back refers to the possible tongue body displacements observable for this vowel pair:

- a) The tongue body can be displaced backwards or
- b) the tongue body can be displaced forward.

In the back configuration, the pharyngeal passage is narrowed relative to the front configuration, and the constriction location is considerably above the glottis. For the front configuration, the constriction location is much lower. Assuming a vocal tract length of about 16 cm, the constriction location for a fronted articulation would be about 4 cm above the glottis, whereas the back articulation would approximately bisect the vocal tract. The fronted articulation yields a maximally high F2, bringing F2 and F3 together. The backed articulation results in a maximally low F2, bringing F1 and F2

together (Stevens 1999). The terms “front” or “back” are, however, misleading if used for the /a/ – vowels, since these terms usually refer to constriction location. It has been already pointed out (Chapter 4.3.1), that the degree of constriction is higher for the “back” /a/ as compared to “front” /a/. Therefore, it is more accurate to distinguish the two vowels, if necessary, by the feature [ $\pm$ constricted]. Moreover, degree of lip aperture is greater for the [+constricted] vowel /a/ (Valaczkai 1998). According to these articulatory configurations the [–constricted] vowel /a/ should expose a lower F1, a higher F2, and a lower F3 than its [+constricted] cognate.

However, in Standard Austrian German, hardly any differences could be found for the two /a/ – vowels with respect to constriction degree, even in the most formal speaking task – the reading of logatomes. Therefore, the spectra in Figures 4.27 and 4.28 are nearly identical, although they are assumed to represent the vowels /a/ (Figure 4.27) and /a/ (Figure 4.28) respectively. There are some differences in bandwidth and amplitude of F3, which might be a consequence of increased acoustic losses due to a lower jaw position for the vowel /a/. But these differences are not consistent, i.e. both /a/ and /a/ can exhibit greater bandwidths for F3.

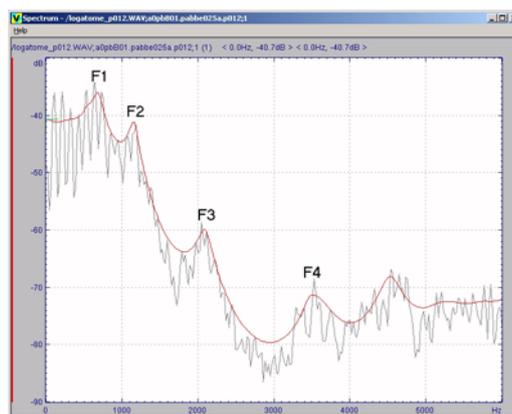


Figure 4.27: Average spectrum of the vowel /a/ taken from the logatome “pabbe”, speaker sp012.

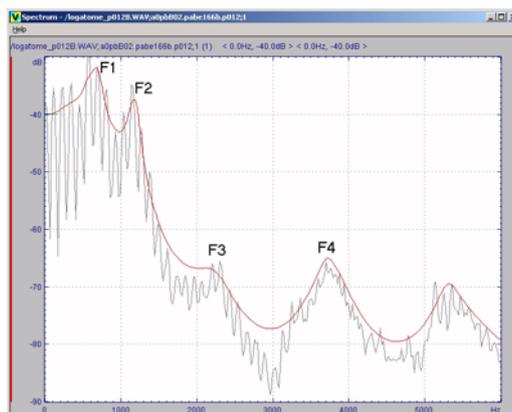


Figure 4.28: Average spectrum of the vowel /a/ taken from the logatome “pabe”, speaker sp012.

Tables 4.33 – 4.35 give the statistical results of the pair-wise comparison of the lower pharyngeal vowels.

F1	L	L	Rs	Rs	Rus	Rus.	Ss	Ss	Sus	Sus
Sp012	<b>641</b>	<b>607</b>	577	597	<b>429</b>	<b>517</b>	547	552	490	458
Sp180	<b>919</b>	<b>827</b>	669	652	<b>434</b>	<b>538</b>	665	639	522	502
Sp082	–	–	623	621	<b>414</b>	<b>518</b>	626	632	<b>477</b>	<b>536</b>
Sp129	–	–	682	641	502	511	<b>665</b>	<b>626</b>	495	514
Sp126	–	–	<b>544</b>	<b>520</b>	<b>368</b>	<b>423</b>	521	486	420	428
Sp127	–	–	665	647	<b>414</b>	<b>476</b>	615	594	429	415

Table 4.33: Mean F1 values of /a/ and /a/ over all speaking tasks. Within each task, the value to the left represents the vowel /a/, the value to the right the vowel /a/. Statistically significant differences within each pair ( $p < 0.05$ ) are marked in bold. Where the t-value is negative, the pair is additionally in italics. Legend as in Table 4.12.

F2	L	L	Rs	Rs	Rus	Rus.	Ss	Ss	Sus	Sus
Sp012	1221	1244	1295	1279	1443	1399	1155	1196	1233	1319
Sp180	1508	1509	1599	1587	1656	1644	1544	1521	1597	1624
Sp082	–	–	1437	1408	1541	1553	1402	1344	1471	1451
Sp129	–	–	<b>1402</b>	<b>1278</b>	1387	1394	1358	1433	1487	1466
Sp126	–	–	1278	1262	1308	1348	1249	1208	1253	1237
Sp127	–	–	1330	1275	1325	1345	1231	1213	1311	1293

Table 4.34: Mean F2 values of /a/ and /a/ over all speaking tasks. Within each task, the value to the left represents the vowel /a/, the value to the right the vowel /a/. Statistically significant differences within each pair ( $p < 0.05$ ) are marked in bold. Where the t-value is positive, the pair is additionally in italics. Legend as in Table 4.12.

F3	L	L	Rs	Rs	Rus	Rus.	Ss	Ss	Sus	Sus
Sp012	2095	2061	2052	2038	2174	2111	2024	2014	2067	2180
Sp180	2417	2412	2450	2505	<b>2783</b>	<b>2592</b>	2511	2602	2662	2614
Sp082	–	–	2402	2421	<b>2544</b>	<b>2448</b>	2508	2536	2520	2513
Sp129	–	–	2457	2391	2538	2519	2390	2328	2473	2433
Sp126	–	–	2095	2055	2206	2214	2129	2164	2189	2263
Sp127	–	–	<b>2407</b>	<b>2305</b>	2590	2511	<b>2463</b>	<b>2327</b>	2506	2462

Table 4.35: Mean F3 values of /a/ and /a/ over all speaking tasks. Within each task, the value to the left represents the vowel /a/, the value to the right the vowel /a/. Statistically significant differences within each pair ( $p < 0.05$ ) are marked in bold. Where the  $t$ -value is negative, the pair is additionally in italics. Legend as in Table 4.12.

As can be read from Tables 4.33 – 4.35, there are hardly any differences between /a/ and /a/. F1 is discerned in the most formal speaking task of reading logatomes and in the stressed position by two speakers, whilst F2 is not discerned at all. F3 is only discerned by two speakers in unstressed positions in the sentence reading task, and by one speaker in the stressed position in reading sentences and spontaneous speech. As concerns constriction location, the two /a/ – vowels are not discerned at all. They are, however, sometimes discerned by either degree of lip opening or degree of constriction.

#### 4.4.6. The role of F0

It is generally regarded as a phonetic universal that F0 correlates with vowel height (Maddieson 1997, Fischer-Jørgensen 1990, Peterson & Barney 1952, Lehiste & Peterson 1961, Neweklowsky 1975, Whalen & Levitt 1995, Whalen et al. 1999), i.e. high vowels expose a higher F0 than low vowels. These results strongly suggest a physiological mechanism which leads to a higher F0 in dependence on the degree of constriction. Numerous studies have been carried out in order to investigate physiological reasons for intrinsic F0 (see Connell 2002 for a concise overview), with the result that still too little is known about the complex mechanism of laryngeal control (Connell 2002: 121). However, several investigations provide evidence for a phonological basis of F0 (e.g. Honda & Fujimura 1991, Kingston & Diehl 1994,

Connell 2002), therefore pointing rather to a planning of F0 than to an automatic mechanism as proposed by Whalen et al. (1998, 1999).

For Standard Austrian German, the traditional division of vowels according to several vowel heights (= tongue height) has been abandoned in favour of contrastive constriction locations with two degrees of constriction at each location. Variations in constriction degree are reserved to variations in speech style rather than to contrastive function. Moreover, the traditional division also lacks physiological realism, since it has been proved in many investigations that the tongue has a higher position for the vowel /a/ than for /ɔ/ or even /o/ (Wood 1987, Bohn et al. 1992, Hoole & Mooshammer 2002).

Nevertheless, at least in stressed positions, the F0 of the vowel /a/ is significantly lower than for the vowels /i/ and /u/, thus corroborating the results of Neweklowsky (1975), who compared the F0 of /a/ versus /i/ and /u/ in stressed positions for Austrian German speakers. This result holds for all speakers and all speaking tasks. However, in unstressed positions, differences between the vowel /a/ and the vowels /i/ and /u/ do not exist for any speakers in any speaking tasks. As concerns the vowels /e/ and /o/, most speakers display differences for the vowel /a/ in stressed positions. Only speaker sp012 has a significantly lower F0 for the vowel /a/ as compared to the vowels /e/ and /o/ in two speaking tasks (reading logatomes and sentences). Speaker sp082 has a significantly higher F0 for the vowel /e/ in spontaneous speech, speaker sp127 has a significantly higher F0 for the vowel /o/ in spontaneous speech, and speaker sp180 has a significantly higher F0 for the vowel /o/ in reading logatomes and reading sentences. In unstressed positions as well, significant results only show up sporadically (speaker sp012: /e/ in the spontaneous speech task, /o/ in the sentence reading task; speaker sp180: /e/ and /o/ in sentence reading task). In other words, no clear picture emerges with respect to a difference between /a/ and the vowels /e, o/<sup>81</sup>. The traditional

---

<sup>81</sup> It has to be emphasized that the vowels /e, o/ and /i, u/ are not distinguished by tongue height, but by constriction location, therefore, /e, o/ should behave in the same way as /i, u/, if the correlation tongue height and F0 were compelling.

assumption that F0 correlates with tongue height can not be corroborated by the data of Standard Austrian German.

It is nevertheless of interest whether F0 plays any role in distinguishing the vowels of Standard Austrian German. If any correlation with the traditional parameter tongue height exists, then the respective unconstricted vowel should expose a lower F0. Such a distinction is only realized by speaker sp012 for some vowel pairs in the task of reading logatomes. Speaker sp180 does not vary F0 according to constriction degree (see Table 4.36).

Logatomes	i – ɪ	y – ʏ	e – ε	ø – œ	u – ʊ	o – ɔ
Sp012	134–132	134–135	<b>120–115</b>	<b>123–119</b>	<b>138–132</b>	<b>123–116</b>
Sp180	206–216	212–225	172–177	184–180	227–229	182–178

Table 4.36: Mean F0 values of all vowels of the task of reading logatomes. Vowels are grouped in pairs, within each pair, the value to the left represents the [+constricted] vowel, the value to the right the [–constricted] vowel. Statistically significant differences within each pair ( $p < 0.05$ ) are marked in bold.

In the task of reading sentences, hardly any statistically significant differences with respect to constriction degree occur in stressed positions (see Table 4.37).

Sentences	i – ɪ	y – ʏ	e – ε	ø – œ	u – ʊ	o – ɔ
Sp012	128–126	151–125	109–111	126–117	140–132	112–106
Sp180	206–218	230–224	173–171	223–200	251–262	179–194
Sp082	226–222	233–229	197–190	<b>213–199</b>	233–218	192–185
Sp129	186–191	196–198	144–151	176–168	197–182	148–157
Sp126	129–127	136–138	113–118	124–123	<b>137–120</b>	113–119
Sp127	165–169	184–173	133–140	157–152	<b>182–160</b>	<i>138–155</i>

Table 4.37: Mean F0 values of all vowels in stressed positions in the sentence reading task. Vowels are grouped in pairs. Within each pair, the value to the left represents the [+constricted] vowel and the value to the right the [–constricted] vowel/. Statistically significant differences within each pair ( $p < 0.05$ ) are marked in bold. Where the t-value is negative, the pair is additionally in italics.

The same picture emerges for stressed positions in spontaneous speech (see Table 4.38):

Spont.	i – ɪ	y – ʏ	e – ε	ø – œ	u – ʊ	o – ɔ
Sp012	108–104	x –122*	90–95	97–87*	106–99*	91–92
Sp180	190–202	185–196	<b>162–139</b>	149*–x	198–181	166–162
Sp082	<b>170–203</b>	234–290*	189–178	197*–x	209–177*	186–178
Sp129	186–178	188–201*	154–153	157–192*	182–181	154–155
Sp126	121–118	x –114*	<b>109–117</b>	x –114*	120–116	110–115
Sp127	123–120	<b>110–131</b>	114–111	113*– x	123–128	119–114

Table 4.37: Mean F0 values of all vowels in stressed positions in spontaneous speech. Vowels are grouped in pairs. Within each pair, the value to the left represents the [+constricted] vowel and the value to the right the [–constricted] vowel. Statistically significant differences within each pair ( $p < 0.05$ ) are marked in bold. Where the t-value is negative, the pair is additionally in italics. The asterisk indicates that no t-tests could be performed.

Tables 4.36 and 4.37 show that statistically significant differences occur only sporadically, if at all. In unstressed positions, the discriminative power of F0 with respect to constriction degree is not any better. Therefore, in Standard Austrian German, F0 does not depend on tongue height.

However, another grouping can be filtered out from these results, and this grouping has to do rather with constriction location rather than with constriction degree. In the front region, the pre-palatal vowels are kept apart from the mid-palatal vowels. In the back region, the pharyngeal vowels are kept apart from the velar vowels. The pharyngeal vowels might be further differentiated, insofar as the /a/ – vowels might show up lower F0 values than the /o/ – vowels.

A statistically significant difference between the upper and the lower pharyngeal vowels (in the sense that the lower pharyngeal vowels expose a lower F0 than the upper pharyngeal vowels) is found in the task of reading logatomes by both speakers, in stressed positions in the task of reading sentences by speakers sp012, sp180, sp127, and in unstressed positions in the task of reading sentences by all speakers. In spontaneous speech, this distinction is displayed in both stressed and unstressed positions by all speakers.

The differentiation of the pre-palatal and the mid-palatal and of the velar and the upper pharyngeal vowels in stressed positions is maintained by most of the speakers, and by some speakers in unstressed positions of the sentence reading task (see Tables 4.38 and 4.39).

i/e vowels	Sp012	Sp180	Sp082	Sp129	Sp126	Sp127
logatomes	<b>133-117</b>	<b>211-175</b>	–	–	–	–
Rs	<b>127-110</b>	<b>212-171</b>	<b>224-193</b>	<b>188-148</b>	<b>128-115</b>	<b>167-137</b>
Rus	<b>117-112</b>	<b>230-216</b>	210-205	<b>186-174</b>	129-128	<b>155-147</b>
Ss	<b>107-93</b>	<b>197-149</b>	194-183	<b>182-153</b>	<b>120-114</b>	<b>122-112</b>
Sus	99-103	<b>191-177</b>	196-204	181-179	<i>117-122</i>	123-123

Table 4.38: Mean F0 values of all /i/ and /e/ vowels of all speaking tasks and all speakers. Vowels are grouped in pairs, within each pair, the value to the left represents the [+pre-palatal] vowel, the value to the right the [–pre-palatal] vowel. Statistically significant differences within each pair ( $p < 0.05$ ) are marked in bold, in case of a negative t-value, the pair is additionally in italics. L = Logatome reading task, Rs = Sentence reading task, stressed vowels, Rus = Sentence reading task, unstressed vowels, Ss = Spontaneous speech, stressed vowels, SuS = Spontaneous speech, unstressed vowels, Sp = Speaker.

u/o vowels	Sp012	Sp180	Sp082	Sp129	Sp126	Sp127
logatomes	<b>135-120</b>	<b>228-180</b>	–	–	–	–
Rs	<b>136-109</b>	<b>156-186</b>	<b>227-189</b>	<b>190-152</b>	<b>130-116</b>	<b>171-145</b>
Rus	<i>106-117</i>	<i>207-223</i>	199-210	174-181	<i>121-131</i>	148-155
Ss	98-92	<b>196-164</b>	209-182	<b>182-154</b>	118-113	<b>125-116</b>
Sus	98-103	185-177	<b>209-188</b>	186-181	121-119	125-124

Table 4.39: Mean F0 values of all /u/ and /o/ vowels of all speaking tasks and all speakers. Vowels are grouped in pairs, within each pair, the value to the left represents the [+velar] vowel, the value to the right the [–velar, –lower pharyngeal] vowel. Statistically significant differences within each pair ( $p < 0.05$ ) are marked in bold, in case of a negative t-value, the pair is additionally in italics. Legend as in Table 4.38.

In spontaneous speech, in stressed positions, the i/e vowels are differentiated better than the u/o vowels, i.e. by five speakers vs. three speakers respectively. In unstressed positions, only one speaker differentiates the respective vowel pairs by F0. Therefore, F0 might act as an additional cue to differentiate the pre-palatal from the mid-palatal

vowels, the upper-pharyngeal from the velar vowels, and the pre-palatal and velar vowels from the lower-pharyngeal vowels in the most formal speaking tasks.

#### 4.4.7. Evaluation of the results

The acoustic analysis clearly corroborates the assumption of five constriction locations for Standard Austrian German. Therefore, the features [ $\pm$  front], [ $\pm$  pre-palatal], [ $\pm$  velar], and [ $\pm$  lower pharyngeal] can be maintained. Lip protrusion was also verified, thus justifying the feature [ $\pm$  round]. As concerns the vowel pairs on each location, the discriminatory power of F1 is higher than that of F2 or F3. In case a vowel pair shows some tendency towards neutralization, F1 is usually the last discriminatory parameter to be discarded. An exception to this rule are the /u/ – vowels, which, for two speakers, maintain an opposition with respect to F3, in this case, as argued in 4.4.3, shifting the opposition to the degree of lip protrusion. Therefore, as far as can be evaluated from the acoustic data, constriction degree plays a minor role and is outmatched by degree of lip aperture, i.e., at each location, the vowel traditionally termed “lax” exposes a higher degree of lip opening<sup>82</sup>. Constriction degree is only relevant in stressed positions. In unstressed positions, this opposition is easily neutralized. The degree of lip opening, however, is often maintained over all prosodic positions. Therefore, the feature [ $\pm$ open] is added. Differences between the /a/ – vowels only appear sporadically, most of the time, this opposition is given up. Therefore, only one /a/ – vowel is assumed to be of relevance in Standard Austrian German. The tendency to neutralize the /i/ – vowels, the /y/ – vowels, and the /u/ – vowels is not yet that advanced to justify discarding the respective feature. It is, however, of relevance to note that in case neutralization occurs, the [+open, –constricted] vowel assimilates to the [–open, +constricted] one. This challenges assumptions that sound changes have anything to do with ease of articulation

---

<sup>82</sup> This result is not in contradiction with the observation that the /e/ – vowels have a higher degree of lip opening than the /i/ – vowels.

(but only if it can be accepted that widening the degree of constriction eases articulation). Table 4.40 presents the feature matrix of the vowels of Standard Austrian German elaborated from the acoustic analysis:

	/i/	/ɪ/	/y/	/ʏ/	/e/	/ɛ/	/ø/	/œ/	/a/	/o/	/ɔ/	/u/	/ʊ/
constricted	+	–	+	–	+	–	+	–	+	+	–	+	–
open	–	+	–	+	–	+	–	+	+	–	+	–	+
round	–	–	+	+	–	–	+	+	–	+	+	+	+
front	+	+	+	+	+	+	+	+	–	–	–	–	–
lower pharyngeal	–	–	–	–	–	–	–	–	+	–	–	–	–
velar	–	–	–	–	–	–	–	–	–	–	–	+	+
pre-palatal	+	+	+	+	–	–	–	–	–	–	–	–	–

Table 4.40: Feature matrix of the vowels assumed for Standard Austrian German.

Although the degree of lip opening is more relevant, the feature [ $\pm$ constricted] is not discarded, since some speaker-specific differences could be observed in how the relevant oppositions are handled. Some speakers maintain a difference in constriction degree (e.g. speaker sp127 discriminates the /e/ – vowels only with respect to F2 in unstressed positions in spontaneous speech). For the /u/ – vowels, lip protrusion might also be of relevance. For the /y/ – vowels, F3 might become dominant, pointing to a difference in larynx height. Therefore, from the most formal speaking task to the most informal speaking task, several phonological processes might apply, and these might as well be speaker-specific.

## 5. Coarticulation

### 5.1. What is coarticulation?

The term "coarticulation" is attributed to Menzerath & de Lacerda (1933) who distinguished "Koartikulation" and "Steuerung"<sup>83</sup>. "Steuerung" means that the articulation of a certain phoneme is determined to a large extent by the following phoneme. "Steuerung" occurs in sequences which Menzerath & de Lacerda define as homorganic, for instance the sequence "afa". "Koartikulation", on the other hand, is the preparation of certain configurations in anticipation of the next phoneme:

"Während des k-Verschlusses (9-10) bewegen sich die dabei unbeteiligten Lippen bereits auseinander – dies bezeichnen wir als "Synkinese" oder "Koartikulation" -, um das darauffolgende a vorzubereiten. Bei u würde das z.B. nicht geschehen können." (Menzerath & de Lacerda 1933: 50)<sup>84</sup>

Menzerath & de Lacerda base their argumentation on the observation that the lips never come to rest during articulation<sup>85</sup>, that contrary to prevailing opinion at that time, no onglide, steady state<sup>86</sup> and offglide can be observed,

"..., sondern eine wunderbare Koartikulation, die darauf begründet ist, daß ein Wort, ein Satz, als Ganzes gewollt und als Gesamtstruktur artikulatorisch nach immer wechselnder Kombination aufgebaut, besser gesagt, "verflochten" wird." (Menzerath & de Lacerda 1933: 52)

This "interweaving" is not in contradiction to the fact that they can, in most instances, determine where a given phoneme starts and where it ends, despite the fact that certain gestures, e.g. lip opening or velar opening, are not synchronous with an observed stop

---

<sup>83</sup> "Steuerung has been translated as "steering" or "control" by Nolan & Kühnert (1999).

<sup>84</sup> /k/, however, is in open environment, i.e. between two /a/-vowels. Nevertheless, their observation is correct insofar as lip opening for /a/ also starts at the end of preceding labial consonants /m/ and /f/.

<sup>85</sup> Their equipment only allowed the observation of lip opening movements. Observation of the lip protrusion movements was not possible.

<sup>86</sup> "Stellung", here translated as "steady state", is meant in the sense that all articulators remain in a certain setting for a certain period of time.

or start in the oral airflow. Theoretically, Menzerath & de Lacerda push the problem of segmentation to the level of phonology ("Lautabgrenzung heißt aber offenbar Abgrenzung der Laute", S. 59, see also Nolan & Kühnert 1999, Tillman & Mansell 1980), whilst practically they use the information of the electro-labiograph (for labial consonant - vowel sequences) or the curve of the oral airflow.

Menzerath & de Lacerda do not differentiate between anticipatory and carry-over coarticulation, this is not crucial to their argumentation of the continuous flow of speech. Yet both directions occur in their material and with intraspeaker variability. For example, in the sequences "máma" and "mamá", spoken in isolation, both final vowels appear as non-nasal, independent of their accent, whereas within a sentence, the same wordfinal á, followed by a velar plosive, is slightly nasalized throughout.

High variability in the quality and timing of nasalization has been observed by many researchers so far. Moll & Daniloff (1971) could show that anticipatory nasalization could appear over two vowels preceding the nasal, even if interrupted by a word-boundary<sup>87</sup>, and that quite often there was as much velar opening during the vowels as during the nasal consonant. On the other hand, velar closure NVC sequences started either during the nasal consonant, during the approach to the vowel, or during the steady-state portion of the vowels (Moll & Daniloff 1971: 681). Speaker-specific differences could also be observed in an utterance final sequence /ni/ (from "money"). One speaker exhibited velar closure for the final vowel, one showed a movement towards closure, though closure was not achieved, and two speakers exhibited no velar closure movement, the velum remaining fully open throughout the vowel (Moll & Daniloff 1971: 682).

Language-specific differences in vowel nasalization have been observed by Clumeck (1976), Solé & Ohala (1991), and Cohn (1993), among others. Clumeck (1976) could show that velar coarticulation differs in temporal extent across the six

---

<sup>87</sup> Contrary to Kozhevnikov & Chistovich (1965), who propose a syllable based approach.

languages analyzed. Solé & Ohala (1991) found a different distribution of temporal patterns of nasalization, in American English the extent of nasalization on the preceding vowel was variable according to duration, while in Spanish it remained constant. Cohn (1993) demonstrated that in Sundanese, nasal airflow patterns have plateau-like shapes similar to those observed in French, whereas English exhibits rather smooth, rapid contours.

For Italian, Farnetani (1986, cited in Farnetani 1997) could show that the spread of nasalization depends also on the quality of the vowel. In the sequence /'ana/ the opening of the velopharyngeal port happens at the acoustic onset of the initial /a/ and lasts until the end of the final /a/. In the sequence /'ini/, only a slight anticipation of velopharyngeal opening could be observed during the initial /i/, whereas the port remained open throughout the final /i/ (Farnetani 1997: 376).

Similar results have been obtained for anticipatory lip protrusion (see Farnetani 1999 for an overview). Daniloff & Moll (1968) observed lip protrusion over four consonants preceding the vowel. Lubker (1981) suggested a maximum time of approximately 600 ms for anticipatory lip protrusion. Vaxelaire et al. (2003) found that in /atu/ and /aku/ sequences, lip rounding and tongue dorsum constriction, both associated with the production of the vowel /u/, traverse the intervocalic consonant and may even affect late configurations of the vowel /a/. In a further extended study, Roy et al. (2003) could show for  $V_1CV_2$  sequences spoken by two French subjects, where  $V_1$  is unrounded,  $V_2$  is rounded and the intervocalic consonant either an alveolar or a velar plosive, that anticipatory jaw, labial and lingual gestures associated with the production of  $V_2$  are all initiated before intervocalic consonant contact and that /u/ anticipatory vocalic gestures have a longer extent in the /a/ context than in the /i/ context. Lubker & Gay (1982) not only found language-specific differences:

"..., in general, speakers of Swedish exhibit more extensive movement toward protrusion, produce more accurate target or goal positions, and either begin the movement toward those positions earlier or in relation to the time available to them than do the speakers of American English in this experiment." (Lubker & Gay 1982: 444),

they also found speaker-specific differences among subjects as concerns time-locked vs. look ahead strategies<sup>88</sup>. Perkell & Matthies (1992), based on previous research, proposed a hybrid model, which was a compromise of the former two. This hybrid model is based on the observation that the lip-protrusion gesture comprises two components, a low-velocity initial phase and a more rapid and prominent second phase<sup>89</sup>. As concerns the end of component 2, which is temporally related to the acoustic onset of /u/, they found considerable variation among their subjects (speakers of American English) in the shape and timing of /u/ protrusion trajectories.

"Subject 1 and 4 appear to always end protrusion near the /u/ onset, whereas this constraint seems to be slightly less rigid for subject 3 and weakest for subject 2, whose normal speaking rate was faster than the other three." (Perkell & Matthies 1992: 2923)

This observation of two components in the lip protrusion gesture touches a further aspect, namely, differentiating coarticulation from processes. A phonological aspect of interpreting differences in lip protrusion has been put forward by Boyce (1990) on her analysis of the "trough" – effect in Turkish and English. Several studies on English, Swedish, Spanish and French (McAllister 1978, Gay 1978, Engstrand 1981, Perkell 1986, all cited in Boyce 1990) report that, for sequences of two rounded vowels with intermediate consonants, EMG recordings show a diminution of rounding during the consonant. In her analysis, Boyce (1990) found clear differences between the languages Turkish and English: Turkish speakers showed a consistent plateaulike pattern of movement, whereas the English speakers exhibited the trough pattern. Boyce concludes:

"The complexity of this interpretation lies in the conclusion that different languages may employ different articulatory strategies. In some sense, this is to be expected, since the combination of phonology, lexicon, and syntax in different languages may impose entirely different challenges to articulatory efficiency. In fact, the hypothesis behind this comparison of Turkish and English was the notion that, in contrast to English, Turkish provides ideal conditions for articulatory look-ahead. [...] The finding that current models of coarticulation are insufficient to account for language diversity indicates how difficult it may be to penetrate to the universal level of speech production." (Boyce 1990: 2593f).

---

<sup>88</sup> See Farnetani (1997) and Farnetani & Recasens (1999) for an overview of coarticulation models.

<sup>89</sup> Similar results have been obtained by Benguerel et al. (1977) and Al-Bamerni & Bladon (1982) for nasalization.

Results of own data may add to the discussion. Standard German, as well as Standard Austrian German, differentiate vowels for lip protrusion. This might put some constraint on anticipatory lip protrusion in order to avoid mixing of the front protruded and unprotruded vowels (see Farnetani 1999, but see Lubker & Gay 1982, Vaxelaire et al. 2003, and Roy et al. 2003 for contradictory results, who observed earlier lip protrusion in their Swedish and French subjects respectively). Lip protrusion, consequently, should not start earlier than the consonant preceding the protruded vowel, but rather at the earliest point in time after consonant release. This has been tested for CV sequences for stressed vowels, sentence reading task<sup>90</sup>, where C is either a lenis or fortis alveolar plosive and V a back protruded vowel /o/ or /u/. The acoustics of a CV sequence where C is an alveolar plosive and V a back protruded vowel demand a steep F2 transition starting (theoretically) at approximately 1800 Hz at consonant release and falling to approximately 800 Hz to 600 Hz for the vowel when lip protrusion has been accomplished<sup>91</sup>. Therefore the point in time at which lip protrusion finishes can be tested from F2 at vowel onset and the following transition into the vowel<sup>92</sup>. In the case where F2 is low at vowel onset and shows no more movement, lip protrusion has taken place already during the time from consonant release until vowel onset (VOT). Where F2 is high at vowel onset, however, lip protrusion starts at vowel onset and F2 shows a gradual fall into the vowel.

---

<sup>90</sup> Stressed vowels from reading sentences tasks have been chosen, because Standard Austrian German shows some tendencies towards neutralization of the fortis/lenis opposition for front plosives (see Moosmüller & Ringen 2004, Moosmüller 1991, 1987). Therefore, hardly any fortis plosives with sufficient VOT duration (> 40 ms) can be found in spontaneous speech or in unstressed positions.

<sup>91</sup> For the back protruded vowels, there is not much difference in the formant frequencies of F1 and F2 for female and male speakers due to compensatory strategies and the fact that the main differences between a male and a female vocal tract appear in the pharynx (see Fant 2004).

<sup>92</sup> In /du/ or /do/ sequences, F2 is the result of the backwards movement of the tongue body and lip protrusion. F2 values are lower for protruded than for less protruded back vowels, which shows that lip protrusion adds quite substantially to the value of F2, a drop of approximately 200 Hz to 400 Hz, depending on constriction location.

As a result, for all persons, but especially for the female speakers, a strong negative correlation could be observed between VOT duration and F2 value at vowel onset: the longer the VOT duration, the lower the value of F2 at vowel onset (see Figure 5.1).

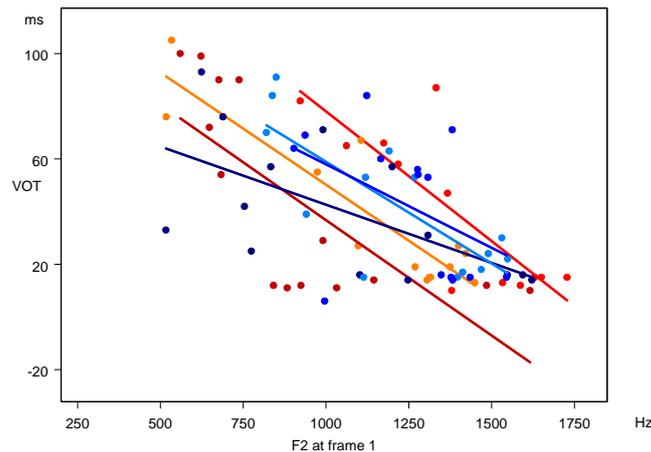


Figure 5.1: VOT vs. F2 at vowel onset for CV sequences, where C is an alveolar plosive and V a back, protruded vowel /u, o/. Reddish lines: female speakers, bluish lines: male speakers.

Table 5.1. gives the PEARSON's correlation coefficient  $r$  for all speakers:

Speaker	sp082	sp129	sp180	sp012	sp126	sp127
$r$	<b>-0.905</b>	<b>-0.735</b>	<b>-0.83</b>	<b>-0.623</b>	<b>-0.788</b>	-0.456

Table 5.1: Correlation coefficient  $r$  for VOT/F2 at vowel onset, broken for all speakers, stressed vowels /u, o/ from the sentence reading task. Statistically significant results ( $p < 0.05$ ) are in bold.

It can be seen from Table 5.1 that, with the exception of speaker sp127, all speakers expose a high, statistically significant negative correlation between the duration of VOT and the value of F2 at vowel onset. This holds especially for the female speakers (sp082, sp129 and sp180), although their overall mean for VOT of fortis plosives does not differ from the male speakers (mean duration of 61.9 ms for the male speakers, 63.5 ms for the female speakers). Therefore, in Standard Austrian German, lip protrusion

starts at plosive release. Whether it is accomplished at vowel onset or not depends on the amount of time that lies between plosive release and vowel onset. These findings are in accordance with results on perception tests which show that in CV sequences, where V is a rounded vowel, lip protrusion has to start at plosive release in order to guarantee the correct perception of the plosive (Maeda 1999). Vaxelaire et al. (2003) found out that, although lip protrusion traverses the preceding plosive in French, the audible part of lip protrusion is located after plosive release, in the VOT span.

In Standard Austrian German, therefore, lip protrusion is fixed with respect to start and duration and does not exhibit the high variability shown for other languages. Mean F2 values and F2 values at vowel offsets of transconsonantal /i/ in /i#CV<sub>2</sub>/, where V<sub>2</sub> is a back vowel, are lower (for some speakers even statistically significantly) than in /i#CV<sub>2</sub>/ sequences, where V<sub>2</sub> is a front vowel (similar results have been obtained by Manuel 1990 and Magen 1997). However, these lower values are triggered by the spectral shape of the intervocalic consonant rather than by lip protrusion, since the spectral shape of the burst of the plosive differs in dependence on the following vowel: the burst spectrum of /Ci/ sequences causes a higher spectral peak associated with F2 than /Cu/ or /Co/ sequences (Fant 1970, see also 5.2.). The following spectrograms will illustrate this interpretation. Figure 5.2 shows the spectrogram of the sequence /i#ti/ from the utterance "die Tiger" (the tigers), Figure 5.3 the spectrogram of the sequence /i#to/ from the utterance "die tote" (the dead: ADJ), for a female speaker.

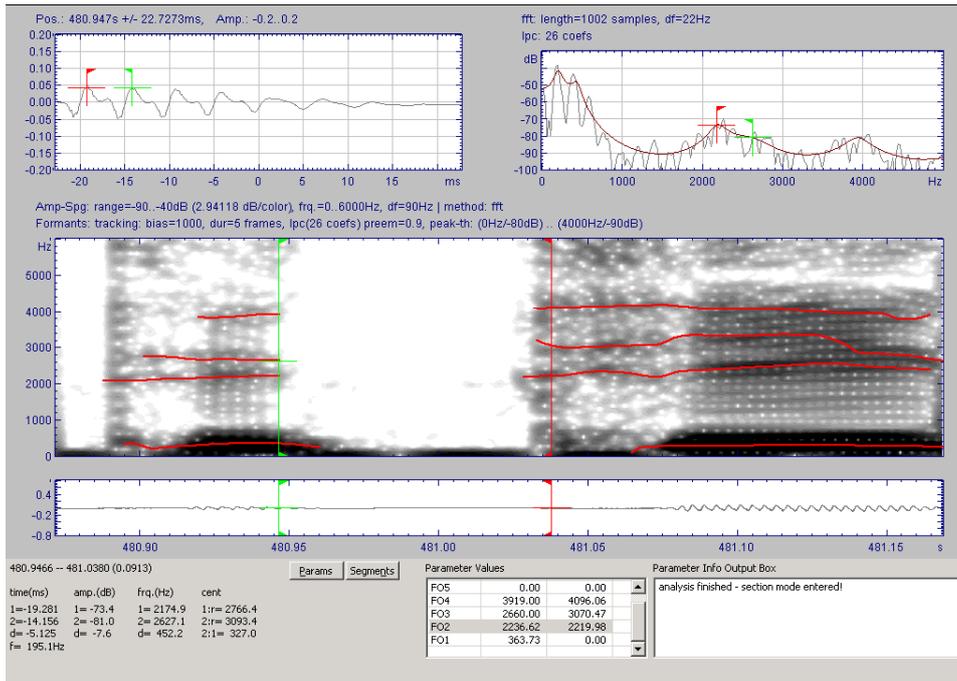


Figure 5.2: Spectrogram of the sequence /di#ti/ from "die Tiger" (the tigers), speaker sp082, sentence reading task. Left cursor positioned at vowel offset, right cursor in the burst of /t/. Values of formant frequencies can be read from the panel below. Bottom panel: waveform window, next panel from bottom: spectrogram window, left upper panel: waveform zoom window, right upper panel: amplitude spectrum window.

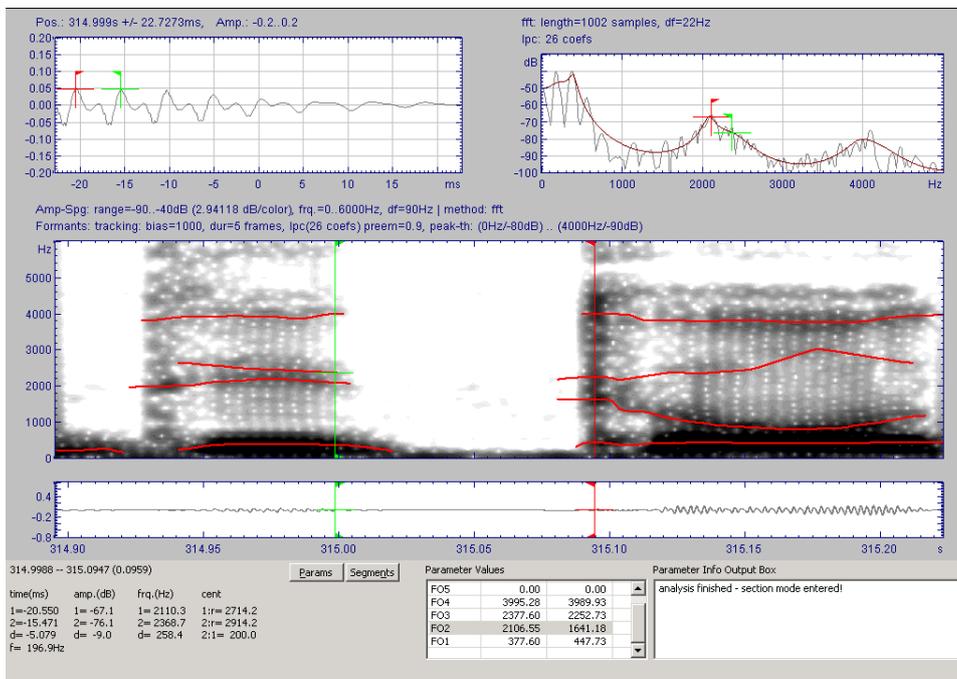


Figure 5.3: Spectrogram of the sequence /di#to/ from "die tote" (the dead: ADJ), speaker sp082, sentence reading task. Left cursor positioned at vowel offset, right cursor in

the burst of /t/. Values of formant frequencies can be read from the panel below. Bottom panel: waveform window, next panel from bottom: spectrogram window, left upper panel: waveform zoom window, right upper panel: amplitude spectrum window.

The burst spectra of the two alveolar plosives differ considerably: In the case of /i#ti/ the tongue is already in a palatal position, causing prominent spectral peaks at approximately 2200 Hz and 3070 Hz, associated with F2 and F3. The formant frequency values at vowel offset of the vowel preceding the plosive, accordingly, have high F2 and F3 values. In the case of /i#to/ (Figure 5.3), two prominent spectral peaks occur at approximately 1650 Hz and 2250 Hz, associated with F2 and F3 respectively. F2 and F3 values at vowel offset are lower than for V<sub>1</sub> in /i#ti/, but still reflect a clear /i/-quality for /i/ in unstressed positions. These lower values point to a smaller contact area of the tongue blade in the case of /ito/ rather than to anticipated lip protrusion. Lip protrusion starts at plosive release and, with a VOT of only 24 ms, the transition is pulled substantially into V<sub>2</sub>.

Transconsonantal anticipatory lip protrusion would trigger much lower F2 and F3 values at vowel offset of V<sub>1</sub> as the ones observed in the /i#to/ sequence discussed for Figure 5.3 or in the other /i#CV<sub>2</sub>/ sequences with C being an alveolar plosive and V<sub>2</sub> a back, protruded vowel. Figure 5.4 exposes an example for transconsonantal anticipatory lip protrusion and tongue body retraction in the sequence /i#hu/ from "die Hupe" (the horn). /h/, being a neutral consonant from an articulatory point of view, would allow a diphthongal movement from /i/ to /u/. It can be seen from Figure 5.4 that F2 and F3 show lower values (1840 Hz and 2250 Hz respectively) at vowel offset, which in turn is already substantially affected by breathiness associated with /h/. However, most of the tongue body retraction and lip protrusion takes place during /h/<sup>93</sup>.

---

<sup>93</sup> V<sub>1</sub>CV<sub>2</sub> sequences, where V<sub>1</sub> and C are not separated by a word boundary, have been investigated in Moosmüller (2007b). Only for one speaker (sp129) transconsonantal lip protrusion could be proved.

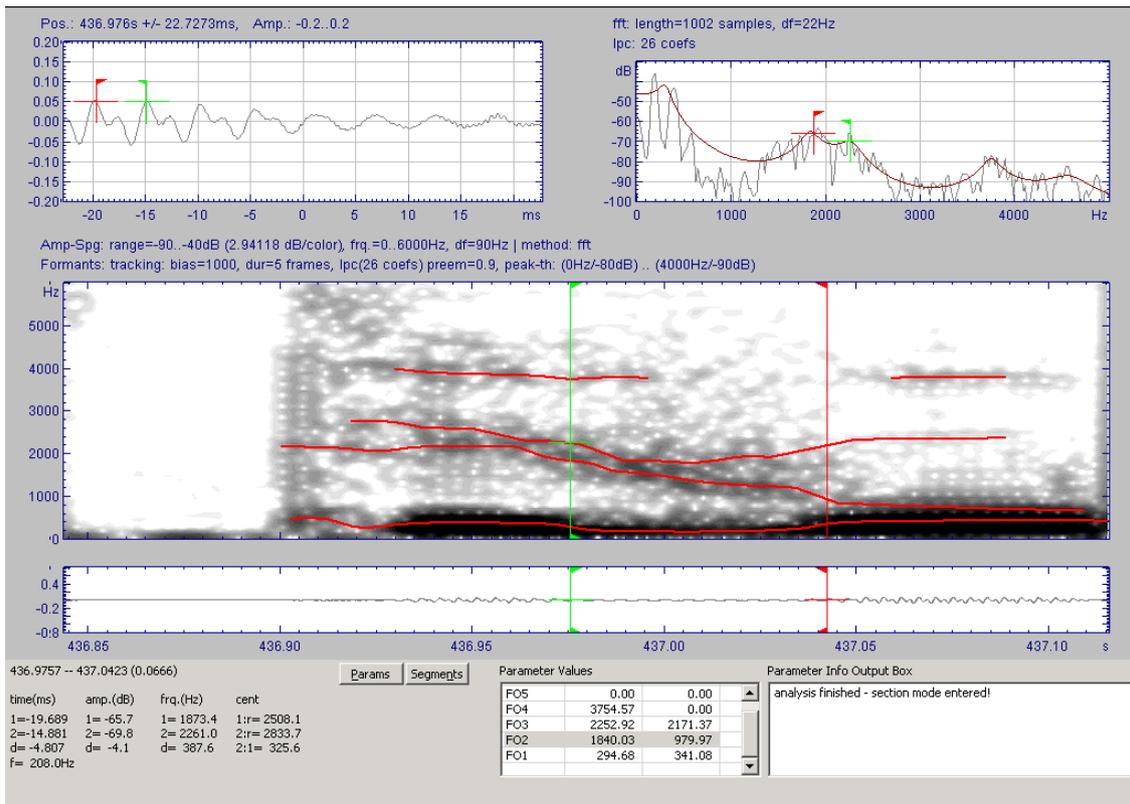


Figure 5.4: Spectrogram of the sequence /di#hu/ from "die Hupe" (the horn), speaker sp082, sentence reading task. Left cursor positioned at vowel offset of /i/, right cursor at vowel onset of /u/. Respective formant frequency values can be read from the panel below. Bottom panel: waveform window, next panel from bottom: spectrogram window, left upper panel: waveform zoom window, right upper panel: amplitude spectrum window.

It can be concluded from these observations of the acoustic signal that, in Standard Austrian German, lip protrusion starts at plosive release and takes some time until it is accomplished<sup>94</sup>. Whether protrusion is accomplished at vowel onset or not, depends on the time between the release of the plosive and the onset of the vowel (VOT). Lower F2 values for transconsonantal /i/ ( $V_1$ ) have to be attributed to the spectral shape (and therefore articulatory configuration) of the intervocalic plosive, which differs according to the quality of  $V_2$ .  $V_2$ , in any of the discussed cases, bears the stress. Therefore, what has been observed in Standard Austrian German, is a sequencing of movements from

<sup>94</sup> In sequences not separated by a word boundary,  $V_1$  of only one speaker (sp129) was affected by lip protrusion (see Moosmüller 2007a).

vowel to consonant to vowel<sup>95</sup> (see also Wood 1996, 1997, Ericsson et al. 1999, Lindblom & Sussman 2004 for the sequencing of articulatory gestures) reflected in the formant frequencies of sonorants or spectral shapes of voiceless obstruents.

The two phenomena which have been chosen to discuss the notion of coarticulation - nasalization and lip protrusion – involve articulators which – from an articulatory point of view - are largely independent from the other articulators and from one another. Therefore, in principle, nasalization and lip protrusion can start at any point in time; whole sequences can be articulated with the velopharyngeal port open or partly open or with protruded lips, phenomena like these can even have a dialectal or idiosyncratic status. The independence of these articulators is reflected in the diversity of the results obtained in the different investigations on these phenomena. The results obtained by Solé & Ohala (1991), Boyce (1990) and by Perkell & Matthies (1992, with previous studies) point to the possibility to interpret these phenomena differently, anticipatory nasalization and lip protrusion being rather an object of the phonology of a language having a processual character than part of coarticulation.

What, then, is coarticulation and how can it be teased apart from phonological or phonetic processes?

Definitions of coarticulation are as diverse as the results of phenomena defined as coarticulation. A very broad view on coarticulation might be attributed to Ohala 1993, who merged assimilation and coarticulation: "Here I will use 'coarticulation' and 'assimilation' as synonyms" (Ohala 1993: 156)<sup>96</sup>. Such a view of coarticulation can also be found in Chafcouloff & Marchal (1999), who, in discussing nasalization, use assimilation and coarticulation as synonyms:

"The study of nasal coarticulatory effects, or to put it differently, the spreading of the nasal feature onto adjacent segments,..." (Chafcouloff & Marchal 1999: 69)

---

<sup>95</sup> Contrary to Öhman (1966), who assumes a diphthongal motion of vowels onto which consonants are superimposed.

<sup>96</sup> However, he uses both terms in this article, and it does not really clarify what is what.

Another widely read view defines coarticulation as overlap of segments in time and space (e.g. Lindblom 1983, Nakamura 2005, Tabain 2001, Farnetani 1997):

"The motor events of a sequence of phonemes overlap in space and time. In pronouncing, for example, [ku], the speaker begins to round and protrude his lips in anticipation of [u] before the release of the tongue closure for [k]. This spatial and temporal overlap of adjacent gestures is a very general phenomenon and can be observed in all languages. The term for it is *coarticulation*." (Lindblom 1983: 220)

"During the movements of different articulators for the production of successive phonetic segments overlap in time and interact with one another. As a consequence, the vocal tract configuration at any point in time is influenced by more than one segment. This is what the term "coarticulation" describes." (Farnetani 1997: 371)

"...that the movements of an articulatory organ overlap in time with those of different articulatory organs or different parts of the same organ. This is called coarticulation, the realisation of which is determined language-specifically." (Nakamura 2005: 1)

This view of coarticulation as segmental overlap is rather broad as well, comprising a vast amount of phenomena in the organization of connected speech, whereas others see coarticulation as the neuromuscular production of movements. The latter conceptualization often sees coarticulation as partly planned (Wood 1996, 1997, Ostry et al. 1996, Dang et al. 2004).

"Coarticulation is a natural phenomenon involved in human speech, which originates from movement planning strategies and from physical interactions among speech articulators." (Dang et al. 2004: 25)

"The sounds of speech may be combined in various ways, and the associated articulator movements may vary as the kinematic context changes. This kinematic variation, known as coarticulation, is one of the most pervasive characteristics of speech production. Some aspects of coarticulation may be centrally planned, whereas others may not be planned but may arise from factors such as muscle mechanics, musculoskeletal geometry, and jaw dynamics." (Ostry et al. 1996: 1570)

Recasens (1999), who defines coarticulation as temporal coproduction of gestures, claims to make a distinction between "articulation proper" and "gestural overlap":

"Moreover, a distinction should probably be made between coarticulation proper and gestural overlap: inspection of articulatory data reveals, for example, that the gesture for a vowel may coarticulate slightly with that for a following consonant before the well-defined onset of the consonantal gesture actually occurs. Unfortunately much of the data reported in the literature have been taken at individual points in time along particular utterances and thus, do not allow disentangling these two notions." (Recasens 1999: 81)

Whatever the definition of coarticulation, the unifying bond of all views is the assumption of a "unit" or "canonical segment" which is modified when executed (Farnetani 1990: 94).

"To recapitulate: the concept of coarticulation entails the hypotheses that at some level speakers make use of a representation in terms of abstract phonological segments, and that there are regular principles governing the articulatory integration of those segments in speech." (Kühnert & Nolan 1999: 9)

Coarticulation, is the

"process bridging the invariant and discrete units of representation to articulation and acoustics." (Farnetani & Recasens 1999: 31)

Attributing coarticulation to a mere processual character neglects, however, the physical and physiological aspects of coarticulation put forward by Kühnert and Nolan (1999):

"..., a single vocal tract has to alter its shape to satisfy the requirements of all the sounds in a sequence. The vocal tract is governed by the laws of physics and the constraints of physiology, but (...) it is producing its communicative artefact in 'real time'. It cannot move instantaneously from one target configuration to the next." (Kühnert & Nolan 1999: 8f).

Therefore, it is those "regular principles" (Kühnert & Nolan 1999: 9) which need further clarification and division. In their investigation of /sʃ/ sequencies, Holst & Nolan (1995) assume a phonological process of assimilation to account for compensatory lengthening in [ʃ], when no spectral characteristics of [s] are visible anymore, i.e. when total assimilation has taken place. Nolan et al. (1996) repeated this investigation applying articulatory methods to account for the critique put forward by Browman (1995). The results show that one speaker applied a phonological assimilation rule, while the other speaker behaved in a way which suggests a deletion of the /s/. However, they could, in both cases, find no evidence of an [s] gesture. The authors conclude:

"Articulatory Phonology already goes impressively far towards accounting for the phenomena of connected speech by providing an implementable description of the process of articulatory blending. It is fully acknowledged by the authors here and in Holst and Nolan (1995) that very many observed outputs of segmental accommodation are appropriately described in gestural terms, that is, as mechanical consequences of the dynamics of the articulators. It would then be extremely elegant and economical if AP were able to go the whole way, and account for all accommodatory phenomena, making Cognitive Phonology redundant, at least in the area of connected speech processes. Elegance and economy are not enough, however, and the scope of AP must be tested by seeking phenomena it cannot

handle. Such a phenomenon appears to be the case of [s] to [ʃ], if the resultant fricative is homogeneously [ʃ]-like, yet longer than a singleton [ʃ]." (Nolan et al. 1996: 135f)

Consequently, phonological processes and coarticulation have to be kept apart, accounting for different phenomena. Such a view is held by Wood (1996, 1997), who offers a definition, which teases the two phenomena apart:

"... that assimilation and coarticulation are the consequences of two distinct processes. Coarticulation is the general local interweaving of articulator gestures that occurs between instantiations of all phonemes, assimilated or not, ensuring smooth motor activity. Assimilation is an arbitrary and conventional reorganization of the timing of articulator gestures of specific phonemes in defined situations." (Wood 1997: 212)

This view restricts coarticulation to the most necessary interaction of neighbouring segments, transsegmental influences are part of assimilation. This rather narrow but convincing view on coarticulation is based on Wood's observations that conflicting demands on an articulator are sequenced rather than blended. Gesture queueing of conflicting gestures, moreover, needs preplanning. Therefore coarticulation is at least partly preplanned in his conceptualization. From his definition follows that many of the investigations on coarticulation belong in fact to levels other than coarticulation, and, more important, that assimilation is processed before coarticulation. Wood (1996, 1997) offers a definition to separate phonology (assimilation) from phonetics (coarticulation), yet, the crucial question is not only which processes have to be attributed to phonology and which phenomena belong to the domain of phonetics, but, moreover, which phenomena, on the phonetic level, have to be conceptualized as processes and which phenomena have to be attributed to the fact that articulators have to move from one phonetic output to the next. This adds a further step to the path from phoneme to phonetic output: phonetic processes, which account for anticipatory articulation that does not touch the phonology of a language. In the same way as phonological processes, phonetic processes are language specific and add to language specific differences of the same phenomenon. To give an example: in Standard Austrian German, as has been shown, lip protrusion starts with plosive release and takes some time until it is accomplished. However, it cannot be traced in the transconsonantal vowel. In French,

on the other hand, lip protrusion could be observed in the transconsonantal vowel (Roy et al 2003), Daniloff & Moll could observe lip protrusion four segments before the influencing vowel. In the same vein as Wood's definition, it is suggested here that in Standard Austrian German, lip protrusion restricted to the interaction with the neighbouring segment, is a coarticulatory phenomenon, whereas anticipatory lip protrusion has rather to be conceptualized as a phonetic process which might or might not occur (see speaker specific differences in Gay & Lubker 1981). The observation of two stages in the production of lip protrusion or nasalization (Perkell & Matthies 1992, Benguerel et al. 1977, Al-Bamerni & Bladon 1982) further supports this argumentation: step 1 can be conceptualized as a phonetic process, step 2 - the necessary part - is coarticulation. Therefore, the path from phoneme to phonetic output, mediated by "regular principles", consists of phonological processes, phonetic processes and coarticulation as the last step which connects the phonemes formed by phonological or phonetic processes. The same phenomenon can, therefore, be either a phonological process, a phonetic process or coarticulation. To which level a given observation belongs, depends on the phonology and phonetics of a language and is, therefore, language specific. Sociolinguistic aspects might further differentiate varieties or speakers.

Before testing the systematicity of coarticulatory phenomena, a further aspect in the articulatory - acoustic relationship has to be examined.

## **5.2. Articulatory vs. acoustic observations**

Although there are many acoustic studies on coarticulation in a broad sense, coarticulation is, as the term suggests, the domain of articulatory phonetics. Tabain (2001), in her review of Hardcastle and Hewlett's (1999) reader on coarticulation, states:

"At first I was surprised at the inclusion of this chapter [on acoustic analysis], since I normally think of coarticulation as an articulatory phenomenon." (Tabain 2001)

The articulators have to fulfill certain requirements to produce a given sequence, and this process of production is the subject of articulatory phonetics. Yet, Tabain's conclusion positions the role of acoustic phonetics:

"Although the chapter may initially seem like a "downer" after so many mainstream articulatory and even theoretical chapters, it is in fact an excellent reminder that the study of coarticulation is ultimately meaningless unless it can be related to an acoustic output, and hence to the perception of the speech signal." (Tabain 2001)

The crucial point in the discussion whether acoustic phonetics can contribute something to the study of coarticulation is, however, the fact that the results of articulatory phonetics and acoustic analyses cannot be directly compared. To give a simple example: in a sequence  $V_1CV_2$ , where  $C$  is a bilabial plosive, the tongue is said to be free to move from  $V_1$  to  $V_2$ , and at the point of release, the tongue is already in the position for  $V_2$ . This is impressively shown by the x-ray motion films produced by Wood (1996): in the sequence /o#pǎ/ from the Bulgarian utterance "deteto xodi po pǎtishtata" ("The child was walking along the path") the uvular tongue body withdrawal of /o/ and the pharyngeal tongue body approach for /ǎ/ happen during the occlusion phase of the bilabial plosive (Wood 1996: 156, Figure 9). Another example is drawn from English, in which the tongue keeps the position of the vowel [æ] and the lips are close for [m], while the velum is raised for [p] in the utterance "camping" (Kent 1983: 68). From an articulatory point of view, the mandible and tongue can work independently from one another to a large extent and two vowels interrupted by a labial consonant rather perform a diphthongal lingual movement. For this reason, some researchers restrict their analyses on vowel to vowel coarticulation on sequences in which the consonant(s) are bilabial (e.g. Manuel 1990, Magen 1997, Cho 2004)<sup>97</sup>.

"The first and second consonants ( $C_1$ ,  $C_2$ ) were always /b/, whose articulation is known to interfere minimally with the vocalic lingual articulation." (Cho 2004: 146)

However, diphthongal tongue movement during a bilabial closure is not undisputed. So-called troughs (deactivation of the tongue musculature during the bilabial closure) in

<sup>97</sup> Acoustic analyses have been performed by Manuel (1990) and Magen (1997). The study performed by Cho (2004) is an articulatory one.

both symmetrical and non-symmetrical vowel contexts were first documented by Houde (1967, cited in Lindblom et al. 2002). Troughs in non-symmetrical V<sub>1</sub>pV<sub>2</sub> sequences were also found in Alfonso & Baer (1982).

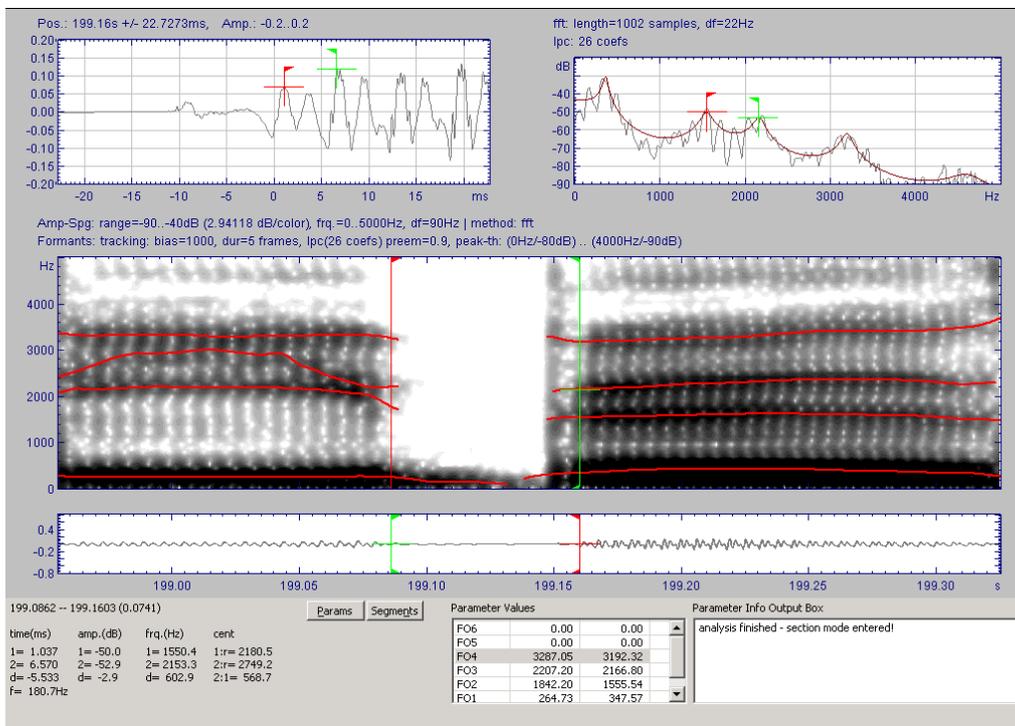
Usually, troughs are investigated in symmetrical vowel contexts. Contrary to the assumption that, during a bilabial closure, the tongue maintains its position in symmetrical vowel contexts, a momentary deactivation of the tongue musculature responsible for the vowel can be observed. Fuchs et al. (2004) could show that troughs are the result of both aerodynamic forces (largest genioglossus posterior deactivation for /p/, smallest for /m/, intermediate for /b/) and a recombination of jaw and tongue movements. The deactivation of genioglossus posterior starts at the beginning of the closure. However, tongue deactivation can also be suppressed; McAllister & Engstrand (1991), could prove language-specific differences:

“In particular, the English and Swedish data both display a deep “trough” in the electrode activation pattern, corresponding to a relaxation of the tongue position roughly coinciding with the consonant; the tendency to such a trough in the French pattern is too weak to be statistically significant.” (McAllister & Engstrand 1991: 9).

For Italian, Farnetani (1991) could show that troughs are minimal when consonants are flanked by /i/, whereas in /a/ contexts, substantial peaks occur during the consonant. The /ipi/ pattern represented in Farnetani (1991) resemble the /ipi/ patterns for French shown in McAllister & Engstrand (1991). These language-specific differences in tongue displacement during a bilabial closure in both symmetrical and non-symmetrical vowel contexts again point out that, where the trough is missing, a process is at work which assimilates the target tongue position of the bilabial plosive to the tongue positions of the flanking vowels (e.g. in French and Italian). Missing troughs, analyzed as phonetic processes, are no counterargument to Lindblom et al.'s (2002) conclusion that the trough effect, in case it is a consistent phonetic phenomenon, gives evidence for a phoneme-by-phoneme execution of articulatory gestures.

Does the trough manifest itself acoustically? Recasens et al. (1997), in analyzing five Catalan speakers, compared EPG data with F2. In /ipi/ sequences, the Qp<sup>98</sup> effects for the bilabial /p/ were very small and short, i.e. hardly any tongue displacement could be observed. F2 and F3, however, exposed a long and large change. These changes in formant frequencies could, since a trough was missing, be attributed to lip closure for /p/ (1997: 550). Vazquez-Alvarez (2005) correlated ultrasound data with F2 in /ipi/ sequences. Although statistically significant tongue body displacements could be proved, correlations with F2 were poor, especially in the C-V<sub>2</sub> sequence. The reason might be that tongue lowering takes place during the closure period; it starts when the closure is complete and at the onset of V<sub>2</sub> the tongue is in its target position. For this reason, neither the offset of V<sub>1</sub> nor the onset of V<sub>2</sub> is affected by a possible tongue lowering.

Therefore, the downward movement of formant frequencies at vowel offset of V<sub>1</sub> preceding a bilabial plosive can be attributed to the closure of the lips (see Figure 5.5).



<sup>98</sup> “Qp stands for the percentage of contact activation over the palatal zone, i.e., number of activated palatal electrodes/total number of palatal electrodes.” (Recasens et al. 1997: 547)

Figure 5.5: Spectrogram of the sequence /ibɛ/ from "Liebe" (love), speaker sp126, sentence reading task. Left cursor positioned at vowel offset, right cursor at vowel onset, respective formant frequency values can be read from the panel below. Bottom panel: waveform window, next panel from bottom: spectrogram window, left upper panel: waveform zoom window, right upper panel: amplitude spectrum window.

The next interpretable event is the burst spectrum, which reflects the configuration of the vocal tract at the point of the release. In case of a bilabial occlusion, the lingual configuration of the vowel following the plosive is finished and, consequently, the spectrum of the plosive reflects this configuration. In Figure 5.5 the vowel following the plosive is [ɛ] and, consequently, a spectral peak can be observed in the region of the F2 of [ɛ] (about 1560 Hz). In Figure 5.6, the following vowel is [ɐ] and, consequently, a spectral peak can be observed in the region of F2 of the a-schwa, namely about 1340 Hz.

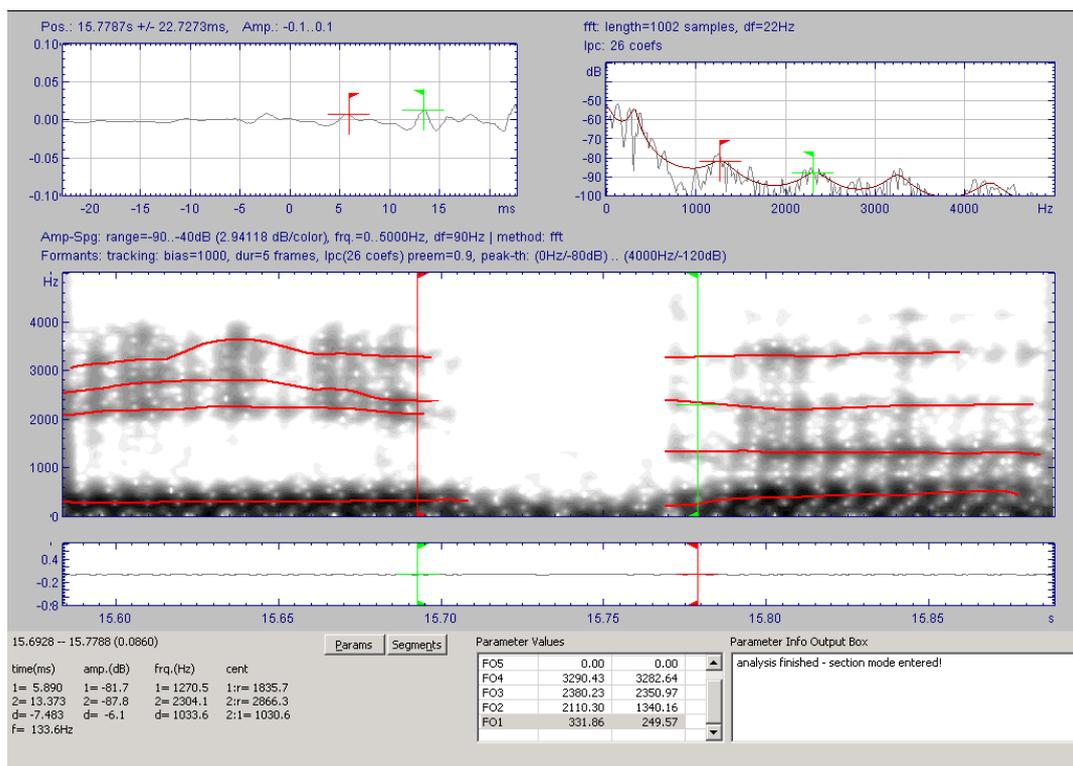


Figure 5.6: Spectrogram of the sequence /ibɐ/ from "lieber" (like something better), speaker sp126, from a repeating sentences task. Left cursor positioned at vowel offset, right cursor at vowel onset. Respective formant frequency values can be read from the panel below. Bottom panel: waveform window, next panel from bottom:

spectrogram window, left upper panel: waveform zoom window, right upper panel: amplitude spectrum window.

In both cases, the /i:/ shows a downward drift of the formant frequencies, which can not be neglected. It should be noted that the /i/s differ in quality; F2 of "Liebe" (love) exceeds 3000 Hz and is located near F4, whereas in "lieber" (like something better), the maximum value of F3 is 2880 Hz. F2 exposes no differences. This difference in vowel quality is attributable to differences in stress (see 6.6.2). This downward drift of formant frequencies is not present in a velar context. Both in the sequence /iɣɛ/ from "winzige" (tiny: PL; Figure 5.7) and in the sequence /i:ɣɛ/ from "Tiger" (tiger; Figure 5.8) F2 and F3 drive towards convergence as the velar constriction or occlusion is reached at the offset of the vowel.

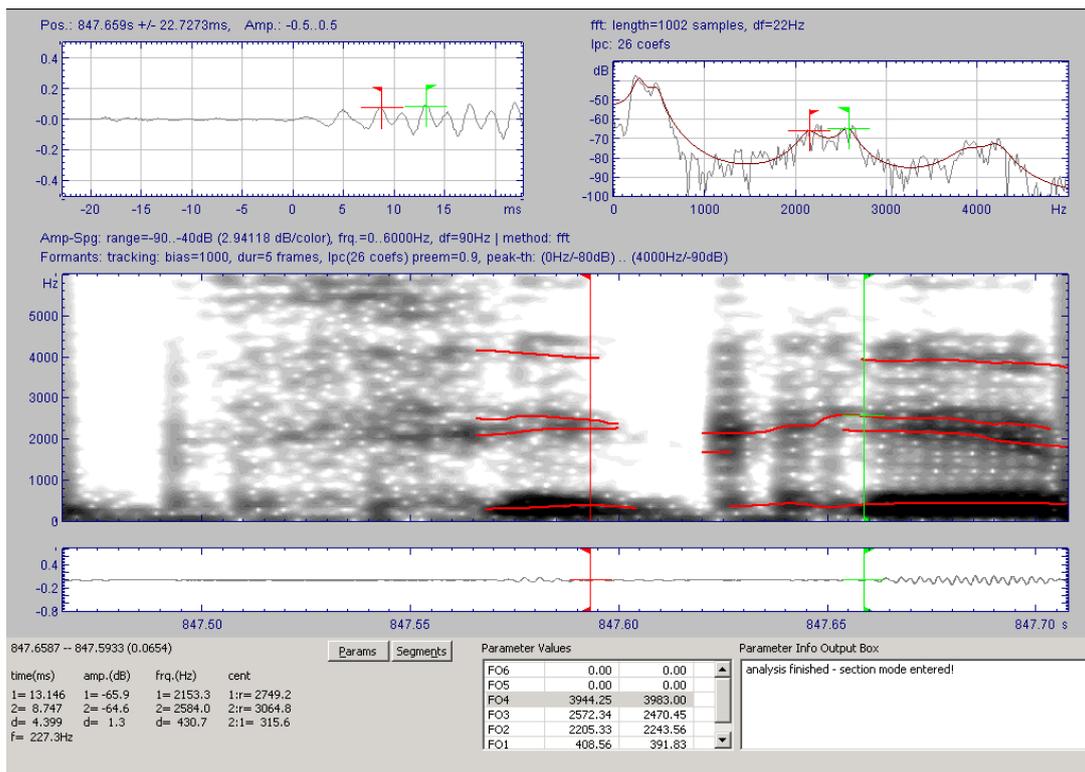
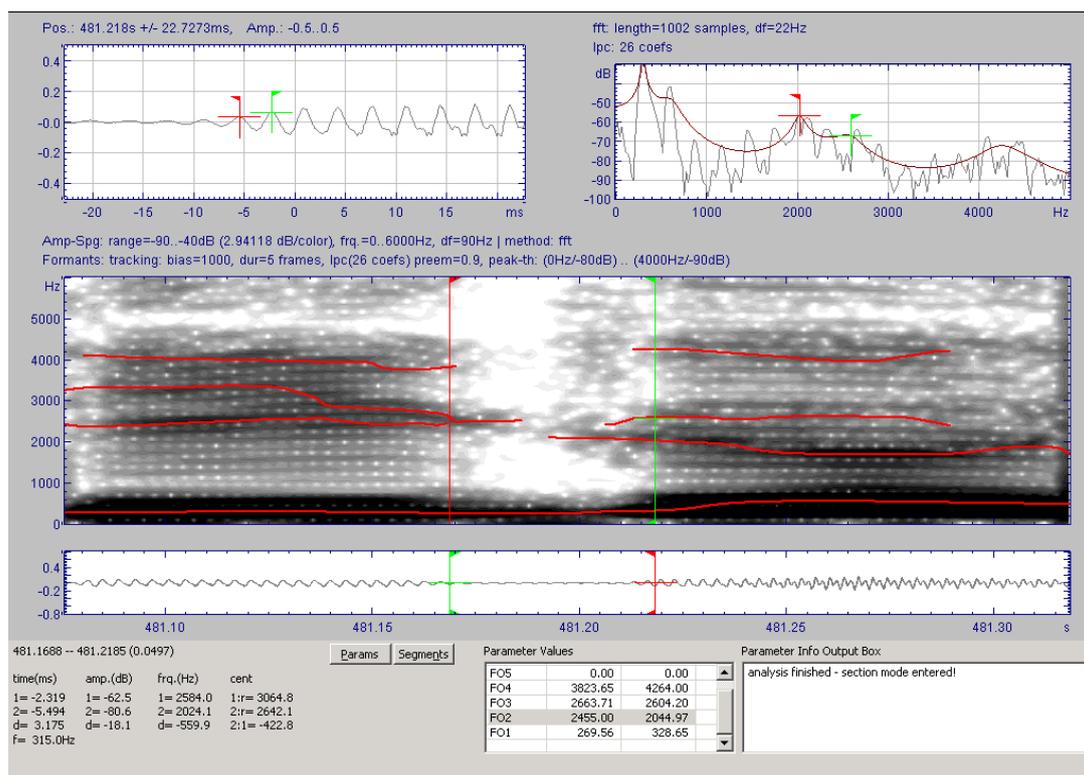


Figure 5.7: Spectrogram of the sequence /iɣɛ/ from "winzige" (tiny: PL), speaker sp082, sentence reading task. Left cursor positioned at vowel offset, right cursor at vowel onset. Respective formant frequency values can be read from the panel below. Bottom panel: waveform window, next panel from bottom: spectrogram window, left upper panel: waveform zoom window, right upper panel: amplitude spectrum window.

The velar plosive of the sequence /iɡɛ/ from "winzige" (tiny: PL; Figure 5.7) exposes multiple releases, and, most interestingly, each individual spectrum has a different shape, indicating a successive re-positioning of the tongue<sup>99</sup>. After the third release, the tongue is in the position for the /ɛ/, and, consequently, no transitional movements of formant frequencies are observable at the onset of the vowel /ɛ/. In Figure 5.8, the approach of the velar constriction – the velar plosive is articulated as a voiced velar fricative – starts shortly after the vowel midpoint at about 50 ms from the vowel onset (total duration of the vowel /i/: 92 ms) and manifests itself in the change in cavity affiliation of F2 and F3 (high F3 and therefore proximity of F3 and F4 in the first part of the vowel, which points to a front cavity affiliation of F3, low F3 and proximity of F2 and F3, pointing to a back cavity affiliation of F3).



<sup>99</sup> Multiple releases can be observed sometimes, especially for velar stop consonants. They are described as a consequence of vibration of the tongue surface that forms the constriction (Stevens 1999: 370). The current observation of changes in the spectral shape would add a phonetic and phonological motivation to such multiple releases, namely positioning the tongue body for the following vowel, which might demand several trials caused by the inertia of the tongue body. Such hypothesis would need to be tested, of course.

Figure 5.8: Spectrogram of the sequence /igʷ/ from "Tiger" (tiger), speaker sp082, sentence reading task. Left cursor positioned at vowel offset, right cursor at vowel onset. Respective formant frequency values can be read from the panel below. Bottom panel: waveform window, next panel from bottom: spectrogram window, left upper panel: waveform zoom window, right upper panel: amplitude spectrum window.

These examples also show that VC and CV sequences are – from an acoustic point of view – not simply mirror images of each other. A comparison of Figure 5.8 and Figure 5.9 demonstrates that the approach of the velar constriction starts shortly after vowel midpoint in the case of /igʷ/ (Figure 5.8), whereas in the case of /gis/ "gieß" (to water: IMP; Figure 5.9), convergence of F2 and F3, typical for velar constrictions or occlusions, is only indicated at the very onset of F2, but at the onset of the vowel (first positive zero crossing), the tongue is already in the position for the vowel, as can be seen by the change in cavity affiliation of F2 and F3, which took place during the time from the burst till vowel onset.

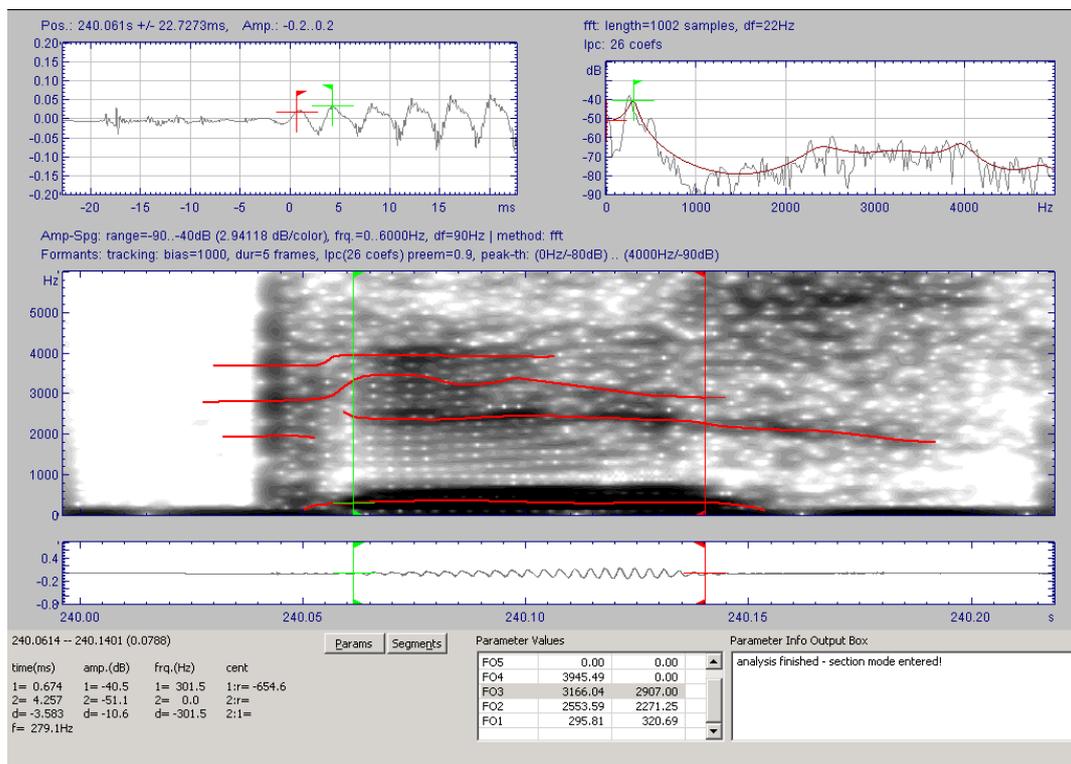


Figure 5.9: Spectrogram of the sequence /gis/ from "gieß" (to water: IMP), speaker sp082, sentence reading task. Left cursor positioned at vowel onset, right cursor at vowel offset. Respective formant frequency values can be read from the panel below. Bottom panel: waveform window, next panel from bottom: spectrogram window,

left upper panel: waveform zoom window, right upper panel: amplitude spectrum window.

The burst spectrum of the velar plosive in "gieß" (to water: IMP) shows two prominent peaks at approximately 1895 Hz and 2756 Hz. This points to a rather fronted constriction location with a front cavity of approximately 4.5 cm<sup>100</sup>. In the time after the burst, the tongue is further fronted to form the constriction for the vowel /i/: at vowel onset the cavity in front of the constriction is approximately 2.8 cm and at the onset of the second period (= 3 ms from vowel onset), the target position is reached, with a front cavity of approximately 2.5 cm.

A similar picture emerges in comparing the acoustic output in VC vs. CV sequences when C is a bilabial plosive. The sequence /ib/ from "Liebe" (love; Figure 5.5) shows a downward drift of F3 in the third part of the vowel (approximately 74 ms from vowel onset; total vowel duration 107 ms), indicating the change in cavity affiliation and preparing the bilabial closure. The downward drift of F2 starts some 10 ms later and is less pronounced.

In the sequence /bi/ from "Bier" (beer; Figure 5.10), formant frequencies have reached their target values for /i/ at the third period from vowel onset, therefore, in the same way as in the /ig/ vs. /gi/ sequences, the transition movement is accomplished quicker and affects the vowel to a less degree in the CV condition as compared to the VC condition.

---

<sup>100</sup> According to Wada et al. (1970, cited after Recasens 1999), velars present as many places of articulation as there are constriction locations for the adjacent vowel. Dembowski et al. (1998) could show that the distribution of the constriction location of the velar plosive "is made up of clusters of smaller distributions corresponding to the individual /k/ allophones" (1998: 37), but see also the discussion on loops in 5.3. Typically, the length of the cavity in front of the constriction is 5 cm to 6 cm (Stevens 1999: 365).

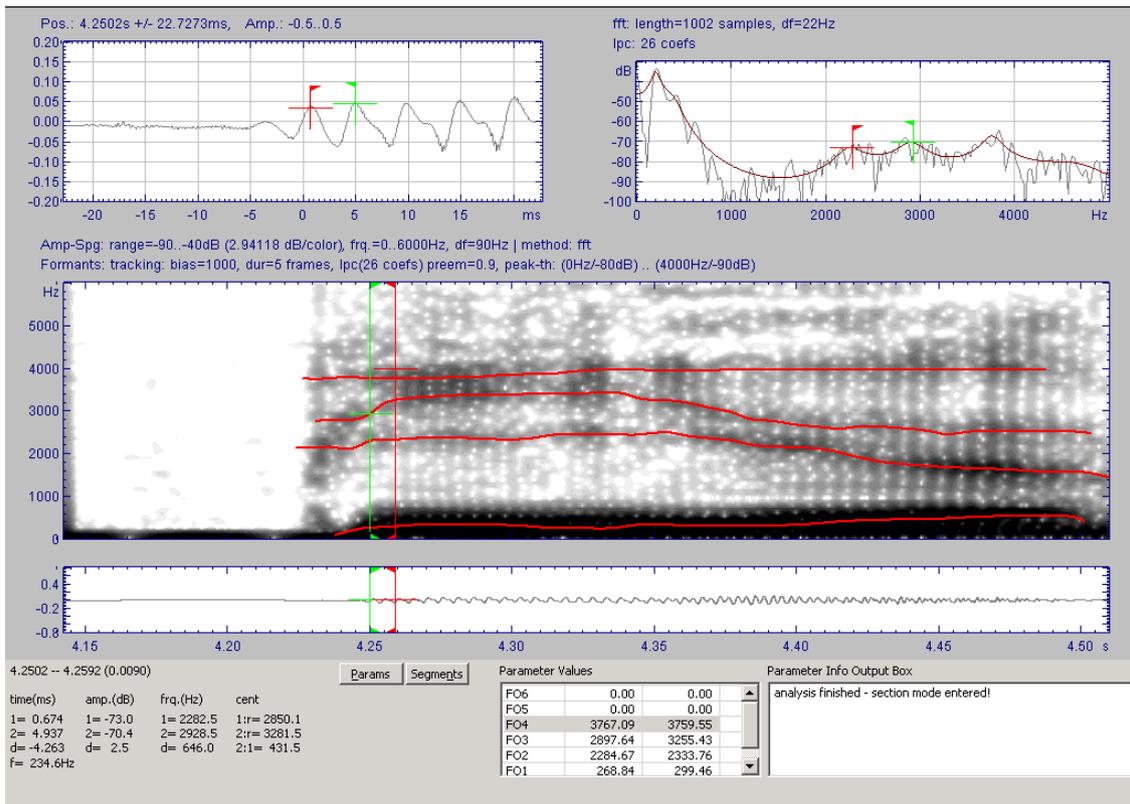


Figure 5.10: Spectrogram of the sequence /bir/ "Bier" (beer), speaker sp082, sentence reading task. Left cursor positioned at vowel onset, right cursor at positive zero crossing of the 3<sup>rd</sup> period. Respective formant frequency values can be read from the panel below. Bottom panel: waveform window, next panel from bottom: spectrogram window, left upper panel: waveform zoom window, right upper panel: amplitude spectrum window.

However, in the bilabial context, it takes the formant frequencies (F2 and F3) longer to reach their target values than in the velar context (3<sup>rd</sup> period in the case of /bi/ and 2<sup>nd</sup> period in the case of /gi/, which, given the difference in F0, amounts to a difference of 6 ms). According to Stevens et al. (1999), there are two components reflected in the formant frequencies for CV sequences, where C is either bilabial or alveolar:

"As the consonant release or closure is formed, there are two components to the movement of the articulators and the formants, particularly for labials and alveolars: one is a rapid movement of the lips or the tongue blade in a time period of 10-20 milliseconds, and the other is a slower movement of the tongue body and mandible toward the following vowel in the case of a release, and from the preceding vowel for a closure." (Stevens et al. 1999: 1117)

In modeling the /bi/ sequence, Manuel & Stevens (1995) calculate a rapid increase of F2 from 1040 Hz to about 1800 Hz as the constriction area of the lips rises from zero to 0.2

cm<sup>2</sup>. Then F2 rises another 200 Hz as the constriction area of the lips increases to about 0.5 cm<sup>2</sup> (Manuel & Stevens 1995: 437). In the present example (Figure 5.10), the calculated F2 of 1040 Hz at constriction area = zero is canceled by a well-developed antiresonance. Two pronounced spectral peaks emerge in the burst spectrum at approximately 2280 Hz and 2900 Hz. Taking into account that the example is spoken by a female speaker, the values measured are in correspondance with the calculations and measurements of Fant (1970) for the Russian palatalised labials.

"The double peak of F2 + F3 of palatalized labials often shows up in spectrograms as an apparent central energy concentration, which might make the labial interval resemble that of palatals. [...] However, the spectrum is distinct from that of palatals owing to the zero just below F2 and the more prominent low frequency region. The two peaks are generally less close than in palatals, and their overall energy is lower. In addition there are, of course, the apparent transitional cues." (Fant 1970: 188)

This means that the tongue is already in the pre-palatal position for /i/ at the time of release; the delay in formant frequency positioning is caused by the second component, the opening of the bilabial occlusion, which takes about 10 ms from the point of release and manifests itself in the upward movement of F2 and F3.

These few examples vividly show that in an acoustic study on coarticulation, the activity of the articulators cannot be seen independently from one other. Moreover, what can be analyzed independently from an articulatory point of view results in one acoustic output containing two or more components. Thus, articulatory and acoustic observations of the same speech events may tell different stories. An x-ray motion film can show what happens during a bilabial occlusion: an inspection of the burst spectrum of a bilabial plosive reveals the vocal tract configuration at the time of release plus the change in area of the lip configuration. What exactly happens during the occlusion (diphthongal movement of the tongue or deactivation of the tongue musculature) cannot directly be observed in the acoustic output.

### 5.3. Processes vs. Coarticulation

The restrictive view on coarticulation put forward in 5.1 implies that coarticulation is systematic and does not vary within a speaker. A given sequence of sounds demands specific adjustments and movements of the articulators involved. If these adjustments and movements are changed, the output is changed as well. E.g. in a CV sequence, where C is an alveolar plosive and V is a back, unrounded vowel /ɑ/, the tongue has to perform a backwards movement to form a constriction in the pharynx; if the tongue is put forward instead, the output will be a front vowel and not a back one. Therefore, there are articulatory movements that are fixed and have to be performed.

From this follows that variations in coarticulations are the consequence of a previously applied process. The output of this process causes different movement adjustments. This argumentation can be proved by an analysis of identical vowels in symmetrical consonant contexts. Figure 5.11 shows the spectrogram of the sequence /divis/ from "Division" (division).

In Figure 5.11, both vowels are in unstressed position. From an articulatory point of view, tongue lowering (trough) might occur during the labial consonant. As concerns the fricative /s/, articulatory studies showed that alveolar fricatives are produced with more tongue grooving and have therefore less contact with the palate (see e.g. Narayanan et al. 1995, Fuchs et al. 2006). Farnetani & Recasens (1993) could show less palatal contact for /i/ caused by alveolar fricatives. Both articulatory adjustments result in a widening of constriction degree and, consequently, in a lowering of F2 and F3. Such a widening of constriction degree takes place in V<sub>2</sub> of the sequence /divis/ (Figure 5.11); although both vowels have phonologically the same quality and although both vowels are unstressed, F2 and F3 of V<sub>2</sub> are substantially lower than F2 and F3 of V<sub>1</sub>.

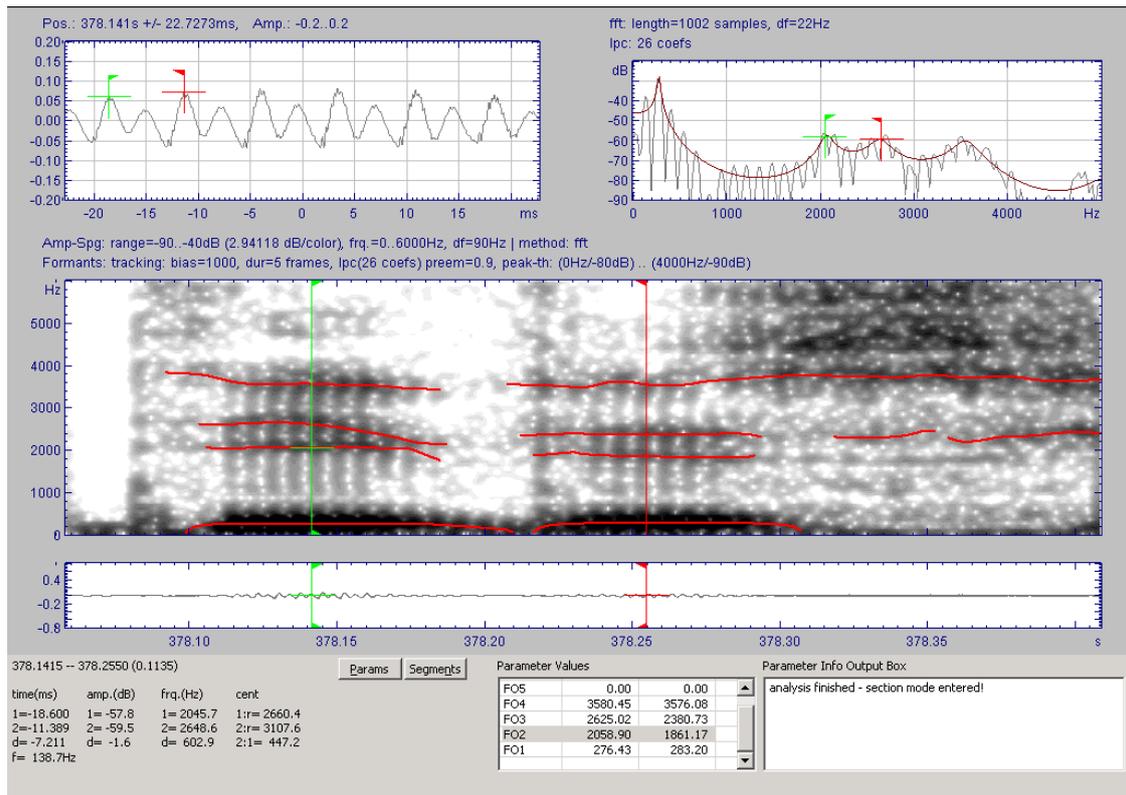


Figure 5.11: Spectrogram of the sequence /divis/ from "Division" (division), speaker sp126, sentence reading task. Left cursor positioned at vowel midpoint of V1, right cursor at vowel midpoint of V2. Respective formant frequency values can be read from the panel below. Bottom panel: waveform window, next panel from bottom: spectrogram window, left upper panel: waveform zoom window, right upper panel: amplitude spectrum window.

However, this is not the case in the sequence /si'fis/ from "Pazifistenstaat" (state of pacifists; Figure 5.12), where V<sub>2</sub> bears the stress. Both F2 and F3 expose higher values for V<sub>2</sub> as compared to V<sub>1</sub>, pointing to a narrower constriction for V<sub>2</sub> as compared to V<sub>1</sub>.

These examples show that, both in /si'fis/ and in /divis/, a process must be at work which changes the quality of V<sub>2</sub> as compared to V<sub>1</sub>, and this process operates before coarticulation connects the phonemes formed by processes. Otherwise, it cannot be explained why, in a phonetic context which demands a widening of constriction degree, F2 and F3 of V<sub>2</sub> rise in the case of / si'fis /.

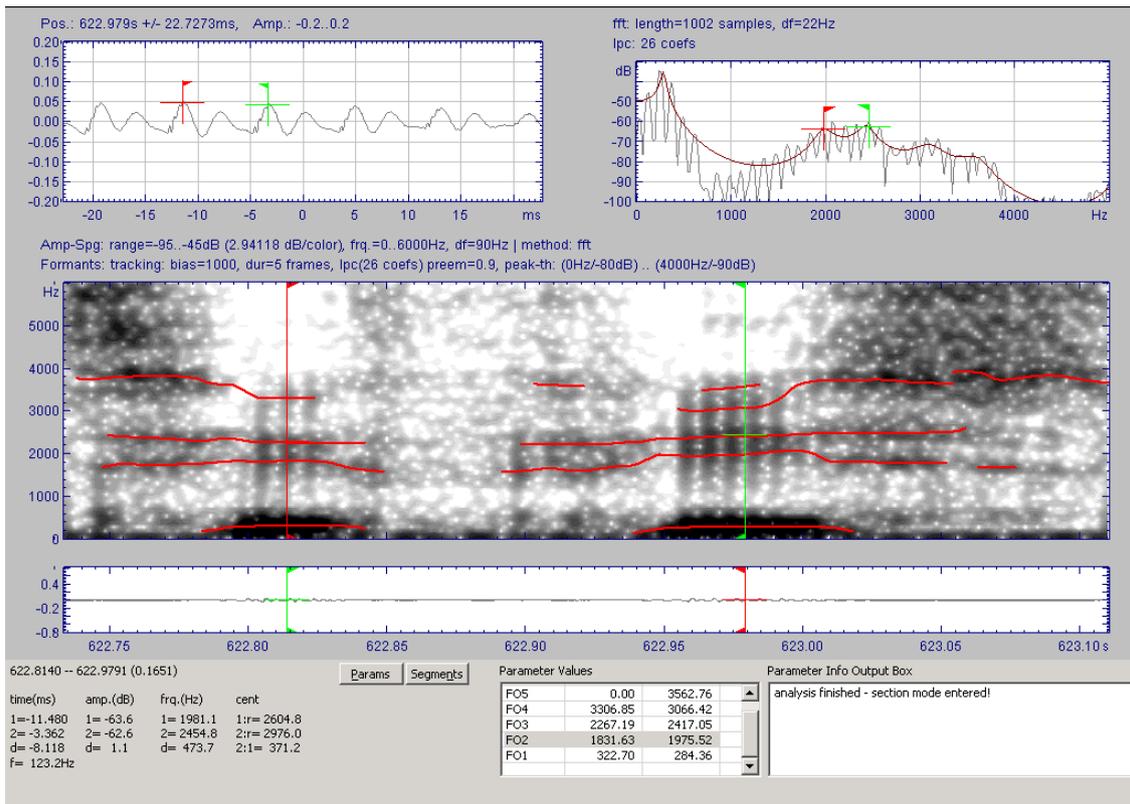


Figure 5.12: Spectrogram of the sequence /sɪ'fis/ from "Pazifistenstaat" (state of pacifists), speaker sp126, sentence reading task. Left cursor positioned at vowel midpoint of  $V_1$ , right cursor at vowel midpoint of  $V_2$ . Respective formant frequency values can be read from the panel below. Bottom panel: waveform window, next panel from bottom: spectrogram window, left upper panel: waveform zoom window, right upper panel: amplitude spectrum window.

Where, then, is coarticulation in the examples of Figure 5.11 and Figure 5.12? In Figure 5.12 /sɪ'fis/, it can be seen that at the very last period of  $V_2$ , a steep fall of  $F_2$  from 2070 Hz to 1830 Hz takes place, i.e. a widening of constriction degree attributable to the following alveolar fricative /s/. This fall is hardly noticeable in Figure 5.11, since the frequency of  $F_2$  is already 1860 Hz at vowel midpoint, i.e., constriction is wide enough already for the following /s/. This pattern is consistent for all subjects of this investigation. I.e., the quality of the vowel varies according to a previous process of – in this case – stress assignment, which affects the degree of constriction. The coarticulation towards the following alveolar fricative, however, varies according to the

previous segment (adjustment of constriction degree vs. no adjustment of constriction degree).

### 5.3.1. The vowel /i/

/s/ shows the least variability among the alveolar consonants (Hoole et al. 1990), i.e. it is quite resistant to coarticulation. The vowel /i/ is attributed high coarticulatory resistance as well (Recasens 1999, Fowler & Brancazio 2000)<sup>101</sup>. Recasens (1999) shows that the palatal contact is higher in the vicinity of /i/ as compared to /a/ and /u/, i.e. consonants are palatalised in the vicinity of a palatal vowel. Palatalization is a process common in many languages (see e.g. Fant 1970 for Russian), and Wood (1996) gives a detailed description of the articulatory movements involved in the palatalization of alveolar consonants in Bulgarian:

"The palatal tongue body gesture of a front vowel in Bulgarian is thus phased in two different ways relative to the occlusion of an adjacent alveolar stop. To palatalize the closure flank the palatal posture of the assimilating vowel is held until the end of the pre-stop vocoid segment before being withdrawn. To palatalize the release flank, the palatal approach of the post-stop vowel is activated already during the preceding vocoid segment and continues during the alveolar occlusion in order to be in place at the release. The two different phasings indicate preplanning of this assimilation." (Wood 1996: 158)

However, palatalization, a natural phonological process, can also be suppressed<sup>102</sup>: Wood (1975a, 1991c) could show that in Swedish palatalization of alveolar consonants is consistently avoided by implementing a pharyngeal movement:

"This pharyngeal maneuver in apical [s] is clearly not a coarticulatory or assimilatory anticipation of a pharyngeal vowel since it occurs in every instance with nonpharyngeal vowels (...), and not when adjacent to a pharyngeal vowel (...)." (Wood 1991c: 288)

Wood (1996) concludes on his Swedish subjects:

"Their alveolar consonants included a pharyngeal tongue body gesture that was antagonistic to the palatal tongue body gesture of an adjacent front vowel, and the two gestures were implemented sequentially rather than coproduced simultaneously." (Wood 1996: 159)

---

<sup>101</sup> The notion of coarticulatory resistance, has, however, to be questioned (see Moosmüller 2007b).

<sup>102</sup> See Stampe (1979), Donegan & Stampe (1979) for the suppression of natural phonological processes.

Öhman's (1966) Swedish subject, however, palatalised the alveolar plosive, so that palatalization or suppression of palatalization might be a dialectal or idiosyncratic feature in Swedish, dependent on whether the plosive is articulated with the tongue tip (apical) or with the tongue blade (laminal) (Wood 1996: 159).

In Standard Austrian German, the burst spectra of plosives are for the most part shaped by the following vowel. The burst spectra of plosives preceding the stressed vowel /i/, therefore, are comparable with the burst spectra of Russian palatalised plosives (Fant 1970). In the same way as for Bulgarian, this palatalization has to be classified as an assimilatory process. Examples for plosives preceding a stressed vowel /i/ are presented in Figure 5.2 (second plosive in the spectrogram exposes /t/ from "Tiger" (tiger)), Figure 5.9 for the sequence /gi/ from "gieß" (to water: IMP), and Figure 5.10 for the sequence /bi/ from "Bier" (beer). In any case it could be shown that the tongue is already in pre-palatal position at release, an observable movement of formants is due to the increase of the lip area (Manuel & Stevens 1995). Unpalatalised /b/ and /d/ would, in any case, expose a sharp rise in formant frequencies at vowel onset (Stevens 1999: 341 and 356), starting at approximately 1050 Hz for the bilabial plosive and at approximately 1600 Hz for the alveolar<sup>103</sup>. In any case, at vowel onset, speakers of the current study expose much higher F2 values than the ones calculated by Stevens (1999) for unpalatalised ones. As concerns the velar plosive<sup>104</sup>, tongue body location is adjusted to the location of the following vowel, i.e. tongue body is fronted before /i/.

Figure 5.13 gives the distribution of the first five frames of F2 and F3, calculated from vowel onset for CV sequences, where C is either bilabial, alveolar or velar and V bears primary stress (task: reading sentences).

---

<sup>103</sup> For a vocal tract length of 17 cm, i.e. an average male speaker.

<sup>104</sup> No modelling for /gi/ is presented in Stevens (1999) or Manuel & Stevens (1995).

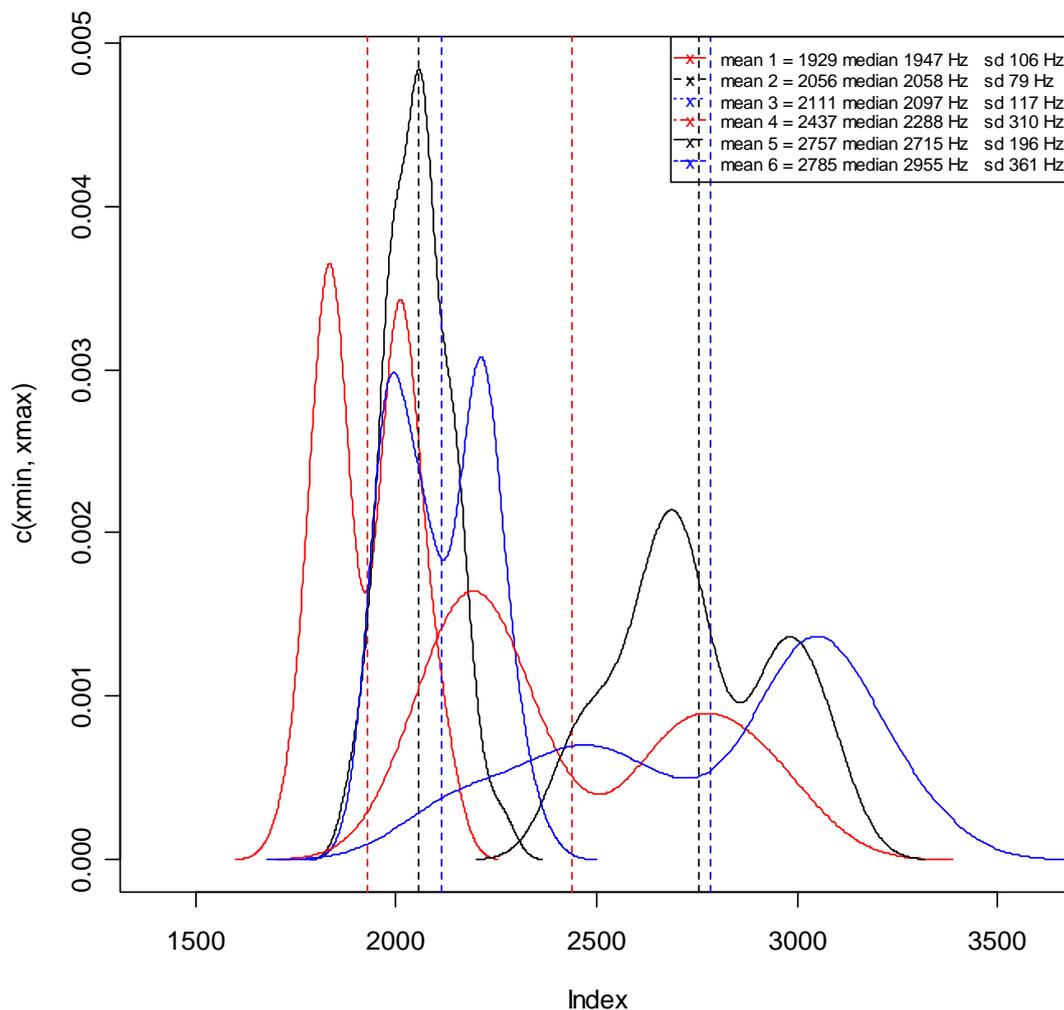


Figure 5.13: Density plot of the mean values of F2 and F3 of the first five frames calculated from vowel onset of all /i/ vowels in stressed position, speaker sp127, sentence reading task. Red lines: C = bilabial, black lines: C = alveolar, blue lines: C = velar.

From Figure 5.13 it can be read that:

- F2 exposes no statistically significant differences with respect to the alveolar and velar context ( $p > 0.05$ )<sup>105</sup>,
- F2 in bilabial and velar context as well as F3 in all contexts exposes two peaks in the distribution,

<sup>105</sup> As can be seen from Figure 5.13, no normal distribution can be assumed. Therefore, a Median Test has been performed.

- F3 shows a high variability in any context.

The /i, y/ vowels in Standard Austrian German are articulated in an acoustically unstable region, this holds especially for /i/, because the front cavity is not lengthened through lip protrusion. As has been stated already, a shift in cavity affiliation takes place when the location of constriction is at about 2/3 length from the glottis. However, as soon as the constriction is widened or constriction length is shortened, this shift in cavity affiliation is withdrawn, F3 is affiliated with the back cavity and consequently drops again. Therefore, /i/ in Standard Austrian German is very sensitive to slight changes in the configurations of constriction degree or constriction length. I.e. in "gib" (to give: IMP) F2 and F3 are both lower than in "gieß" (to water: IMP), consequently, a narrower and longer constriction is planned in "gieß" as compared to "gib". In Figure 5.14 the movements over time of F1, F2 and F3 are plotted on top of each other for two instances of /gib/ and two instances of /gis/, no time alignment has been performed.

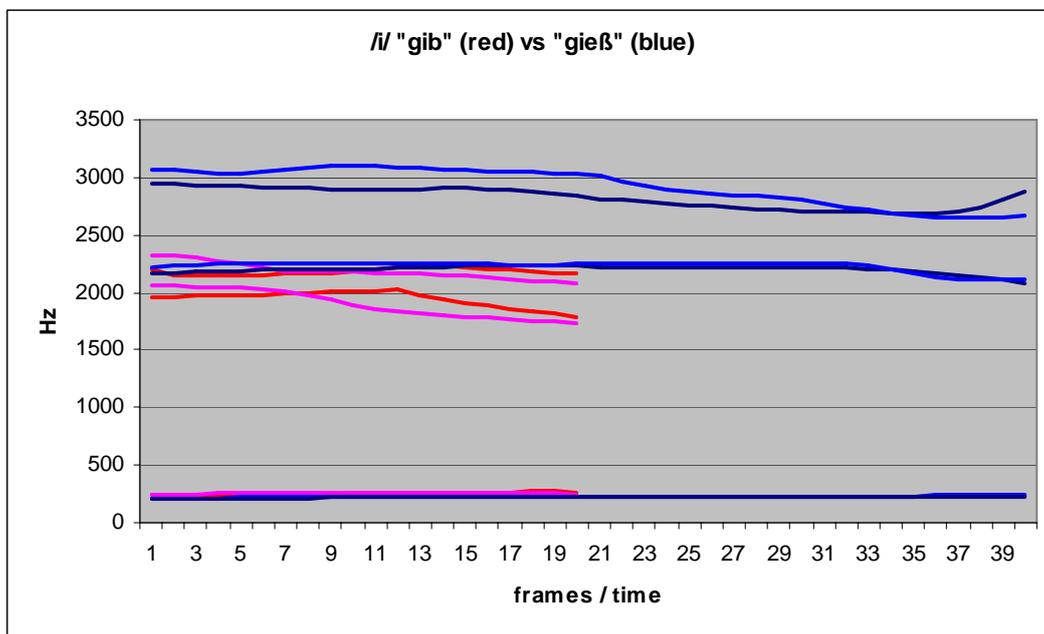


Figure 5.14: F1, F2, F3 movement over time of 2 items of /gib/ "gib" (to give: IMP) and of two items of /gis/ "gieß" (to water: IMP), speaker sp127, sentence reading task.

From Figure 5.14 it can be seen that

- /i/ of "gib" is substantially shorter than /i/ of "gieß",
- F3 of "gib" is substantially lower than F3 of "gieß", overlapping with F2 of "gieß", and
- F2 of "gib" is also lower than F2 of "gieß".

In the case of "gieß", shift in cavity affiliation has taken place, in "gib", F2 is affiliated with the front cavity, F3 with the back cavity. Therefore, although "gib" and "gieß" have the same prosodic condition (sentence initially, primary stress), "gib" is substantially shorter and has a wider constriction degree and/or shorter constriction length than "gieß". This difference in quality does not depend on vowel duration, since, over all instances of /i/, except for p127, no correlation could be found between vowel duration (= number of periods) and either F3 or F2.

However, irrespective of cavity affiliation, in the stressed position the steep and rapid transition movement described for a transition of an unpalatalised plosive towards a palatal vowel could not be observed in the data, i.e. in any case the tongue was in the palatal position at vowel onset, indicating that the tongue had been positioned at an earlier point in time, either in the occlusion phase or after release. Tables 5.2 and 5.3 give the mean values of F2 and F3 of the first five frames from vowel onset in CV sequences, where C is either a bilabial, an alveolar or a velar plosive:

Median F2	/bi, pi/	/di, ti/	/gi, ki/
sp180	2192	2566	2519
sp129	2189	2322	2368
sp082	2225	2324	2318
sp012	1998	2012	2066
sp126	2104	2078	2120
sp127	1947	2058	2097

Table 5.2: Median F2 values of the first five frames calculated from vowel onset in bilabial, alveolar and velar context, stressed position, sentence reading task.

Median F3	/bi, pi/	/di, ti/	/gi, ki/
sp180	2854	3338	3247
sp129	2855	3108	3342
sp082	2670	2999	2989
sp012	2879	3093	3212
sp126	2692	2778	2854
sp127	2288	2715	2955

Table 5.3.: Median F3 values of the first five frames calculated from vowel onset in bilabial, alveolar and velar context, stressed position, sentence reading task.

Since no normal distribution can be assumed for the data, a Median Test has been performed. Two speakers (sp126 and sp082) did not differentiate the alveolar and velar context for both F2 and F3, a further speaker (sp127) did not differentiate the alveolar and velar context for F2. The bilabial and velar context was not differentiated by four speakers (sp126, sp012, sp082, and sp129) for F2, but all speakers differentiated F3 for this context pair. With the exception of speaker sp126, who did not differentiate bilabial and velar context for F2, all speakers differentiated bilabial and velar context. Therefore, the bilabial context is differentiated by all speakers, at least with respect to F3. Differentiation of alveolar and velar context is less secure.

For each subject, F2<sup>106</sup> and F3 are lower in bilabial context. These differences hold also for stressed positions in spontaneous speech. From an articulatory point of view, the lower values for F2 and F3 point out that in many instances, there is less contact between the tongue and the palate when /i/ is preceded by a bilabial plosive than when preceded by an alveolar or a velar plosive. Thus, alveolar and velar plosives increase contact area and are more subject to palatalization than a bilabial plosive.

When, then, does palatalization start? Wood (1996) could observe for Bulgarian that palatalization of the alveolar consonant started already during the vowel preceding the plosive. In order to test whether palatalization starts that early in Standard Austrian German, C<sub>1</sub>V<sub>1</sub>C<sub>2</sub>V<sub>2</sub> sequences were chosen where C<sub>1</sub> is an alveolar plosive /d/, V<sub>1</sub> has a

<sup>106</sup> Except for speaker sp126, who, however, exposes no statistically significant differences for any context, i. e. consonantal contexts are completely blurred.

velar constriction location<sup>107</sup>, C<sub>2</sub> is either labial, alveolar, alveo-palatal, or velar, and V<sub>2</sub> is either a back or a front vowel and has, in any case, a stronger prosodic position than V<sub>1</sub>. Examples are:

"du bald" (you soon): C<sub>1</sub> = alveolar, V<sub>1</sub> = velar, C<sub>2</sub> = labial, V<sub>2</sub> = back

"du sehen" (you see: INF): C<sub>1</sub> = alveolar, V<sub>1</sub> = velar, C<sub>2</sub> = alveolar, V<sub>2</sub> = front

"du kaum" (you hardly) C<sub>1</sub> = alveolar, V<sub>1</sub> = velar, C<sub>2</sub> = velar, V<sub>2</sub> = back

In the case where palatalization would start at the offset of V<sub>1</sub>, F2 of the velar vowel /u/ would substantially rise as the tongue moves from a velar to a palatal location. Figure 5.15 gives an example of four utterances of a male speaker (sp126) where C<sub>2</sub> is a bilabial and V<sub>2</sub> is either a back vowel or a front vowel.

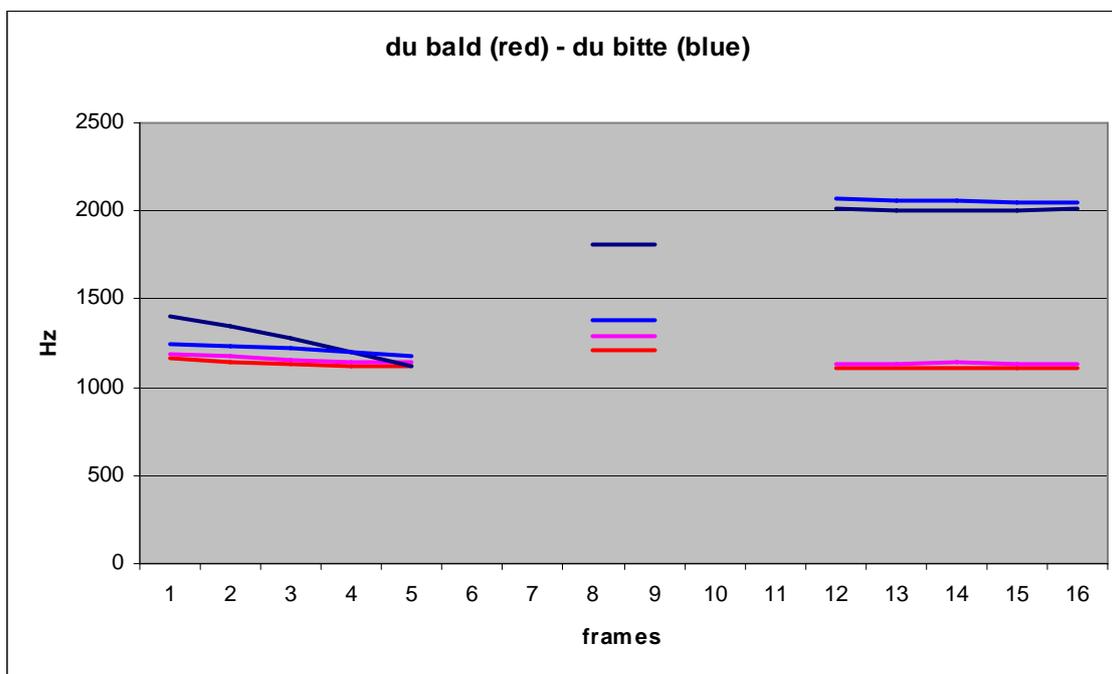


Figure 5.15: F2 movement over time of 2 items of /dubald/ "du bald" (you soon) and of two items of /dubitɛ/ "du bitte" (you please), speaker sp127, sentence reading task. Frames 1 to 5: F2 of the last five frames of /u/, frames 8-9: spectral peaks of the burst of the bilabial plosive associated with F2, frames 12 to 16: F2 of the first five frames of /a/ (red) or of /i/ (blue).

<sup>107</sup> A back vowel has been chosen as the starting point in order to be able to observe whether the tongue starts a movement towards a fronted configuration already at the offset of the back vowel.

Figure 5.15 shows the movements of the last five frames of F2 before the vowel offset of /u/, the spectral peaks associated with F2 of the bilabial plosive and the movements of the first five frames of F2 of V<sub>2</sub>, which is either a back vowel /ɑ/ or a front vowel /ɪ/. It can be seen that the offset of /u/ shows no traces of palatalization. The spectral peaks of the bursts show low F2 values (1206 and 1292 Hz) when followed by a back vowel, the burst spectra of the bilabial plosive differ, however, when followed by a front vowel. In one case the peak associated with F2 is as low as when followed by a back vowel (1378 Hz); in the other case the spectral peak is much higher (1809 Hz). It has to be added that in the case where the spectral peak of the plosive is low, VOT is slightly longer than in the case with a high value (18 ms vs. 13 ms). These two instances of the sequence /du'bi/ show that the tongue is put into position either at plosive release or in the occlusion phase before. Moreover, they also point to the role of VOT: a longer VOT gives the possibility to configure the vowel after release. From the acoustic inspection, however, the transconsonantal vowel V<sub>1</sub> is not palatalised in any instance.

The last five frames of F2 before vowel offset have been submitted to a one-tailed t-Test in order to test whether /u/ is palatalised when followed by a transconsonantal front vowel. In the case of palatalization, the final frames of F2 should show a substantial rise when the tongue is moved from a velar to a palatal location. Table 5.4 gives the mean values of F2 for the final five frames followed by a transconsonantal back or front vowel and the results of the t-Test:

F2 /u/	V <sub>2</sub> [+back]	V <sub>2</sub> [+front]	t	P
sp180	1520.421	1490.076	0.75	0.22
sp129	1233.893	1295.815	1.2	0.12
sp082	1363.959	1379.48	0.52	0.30
sp012	1214.358	1186.268	0.64	0.26
sp126	1180.692	1178.963	0.07	0.47
sp127	1272.421	1298.757	2.42	0.01

Table 5.4: Mean F2 values of the last five frames calculated before vowel offset of /u/ followed by transconsonantal V<sub>2</sub>, which is either a front or a back vowel, sentence reading task. t value and p from one-tailed t-Test.

It can be seen from Table 5.4 that, with the exception of one speaker (sp127), results are not statistically significant. However, the difference of 26 Hz exhibited in sp127 is of no relevance for the current question. It can be concluded that, from acoustic inspection, in the observed set of /du # C<sub>2</sub>V<sub>2</sub>/ sequences, no traces of palatalization could be observed in the movement of F2 of the analyzed last five frames of /u/. Palatalization starts, therefore, either during the occlusion or after the release of the plosive. Lingual configuration of the vowel /i/ is finished at vowel onset. Slight movements of formant frequencies can be attributed to the plosion of the occlusion and subsequent opening of the occlusion area. The results on /Ci/ sequences, where C is either a bilabial, an alveolar, or a velar plosive are summarized as follows:

- At vowel onset, F2 and F3 are already in position for the vowel; no sharp and rapid transition movements can be observed at vowel onset.
- Consequently, a palatal tongue configuration has been adopted before vowel onset.
- This palatal configuration, however, does not reach into the transconsonantal vowel /u/.
- This restricts the time for adopting a palatal configuration either to the occlusion phase or the time after plosive release (VOT).
- /i/ preceded by an alveolar or a velar plosive shows higher F2 and F3 values than /i/ preceded by a bilabial plosive. This points to a larger contact area of /i/ (either a longer or a narrower constriction) in alveolar or velar context. Therefore, lingual plosives actively contribute to palatalization.
- The high variability (especially of F3) is attributable to the acoustically instable constriction location for /i/ in Standard Austrian German.

Coarticulatory effects of /Ci/ sequences would appear as sharp and rapid rises of F2 at vowel onset. These sharp and rapid rises are prevented by palatalization, i.e. by

increasing the contact area between tongue and palate at an earlier point in time. By applying the process of palatalization, coarticulation (e.g. sharp transitions) is smoothed away.

### 5.3.2. The vowel /a/

The vowel /a/ is usually attributed less coarticulatory resistance than the vowel /i/, i.e. is affected to a larger extent by its surroundings than a palatal vowel (see e.g. Farnetani & Recasens 1993). As has been stated in Chapter 4, in Standard Austrian German /a/ is articulated as a back vowel, which means that the tongue body is displaced back and produces a constriction in the pharynx considerably above the larynx<sup>108</sup> and that the tongue tip is displaced back from the inner surfaces of the lower incisors (Stevens 1999: 274). This articulatory configuration results in a low second formant frequency.

Preceded by a bilabial plosive, tongue body configuration can be accomplished at the point of release, i.e. hardly any transition would be visible for F2. This is enforced by the fact that F2, as back cavity resonance, is hardly affected by changes in the front cavity (Manuel & Stevens 1995 and Stevens 1999). Therefore, the increase of the lip area is hardly visible in the movement of F2, but rather appears in the steep and rapid rise of F1.

However, quite often, a clear falling F2 pattern could be observed in /ba/ sequences. Such an example is given in Figure 5.16 which shows the spectrogram of the word [bʌkŋ] "backen" (to bake).

---

<sup>108</sup> As compared to the front vowel /a/, whose constriction location is nearer to the larynx, resulting, consequently, in a higher F2.

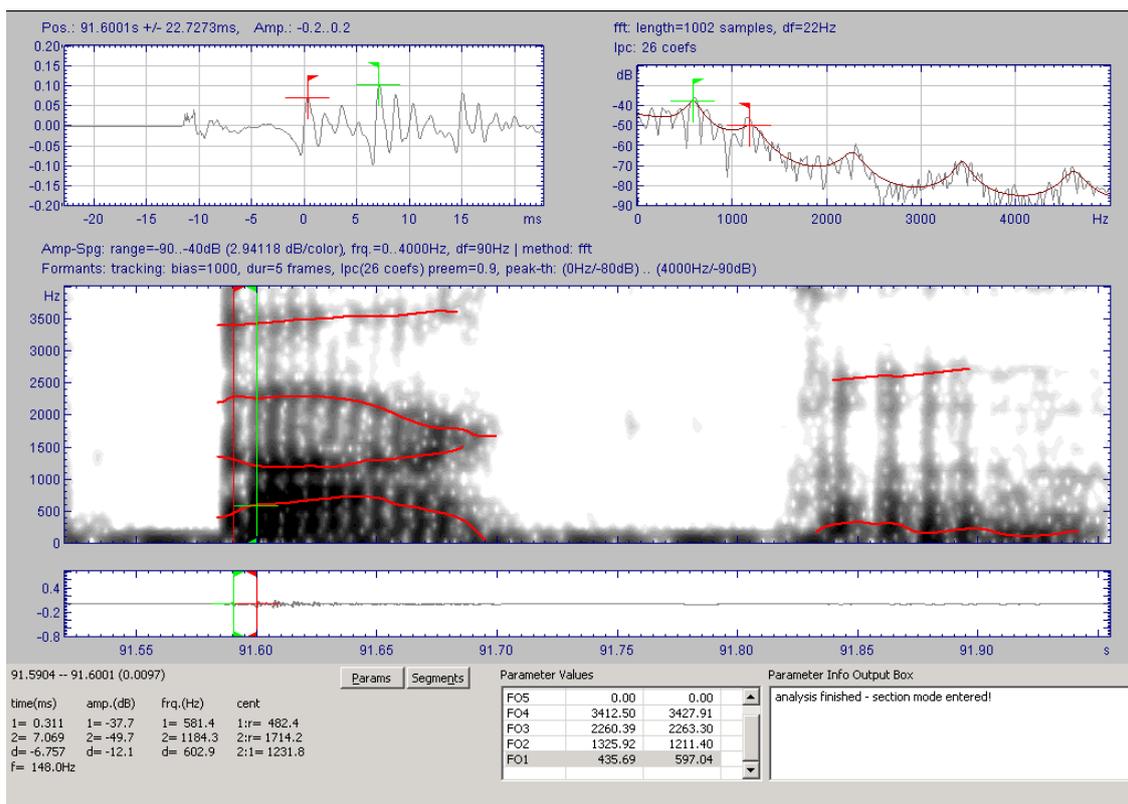


Figure 5.16: Spectrogram of the sequence /'bakŋ/ from "backen" (to bake), speaker sp127, sentence reading task. Left cursor positioned during the burst of the bilabial plosive, right cursor at vowel onset of /a/. Respective formant frequency values can be read from the panel below. Bottom panel: waveform window, next panel from bottom: spectrogram window, left upper panel: waveform zoom window, right upper panel: amplitude spectrum window.

The left cursor in the spectrogram window is positioned in the burst of /b/, the right cursor at vowel onset (1<sup>st</sup> positive zero crossing). It can be seen from spectrogram inspection that F2 falls from release until vowel onset, but at vowel onset, F2 does not change any more until about vowel midpoint. F2 then changes again in preparation of the velar closure. This falling pattern of F2 probably indicates that the tongue body is not yet in position at plosive release. Therefore, as concerns the movement of F2, two /Ca/ – patterns, where C is a bilabial plosive, can be observed: either a falling or a flat pattern. In each case, F2 has its final position at vowel onset.

The production of an alveolar closure is facilitated if the tongue body is placed in a somewhat fronted position and if the mandible is raised (Stevens 1999: 324). When

followed by a back vowel, the tongue body has to move back after release to form a constriction in the pharyngeal region. This backward movement results in a decrease in F2 and a constant increase in F1 (see Figure 5.17).

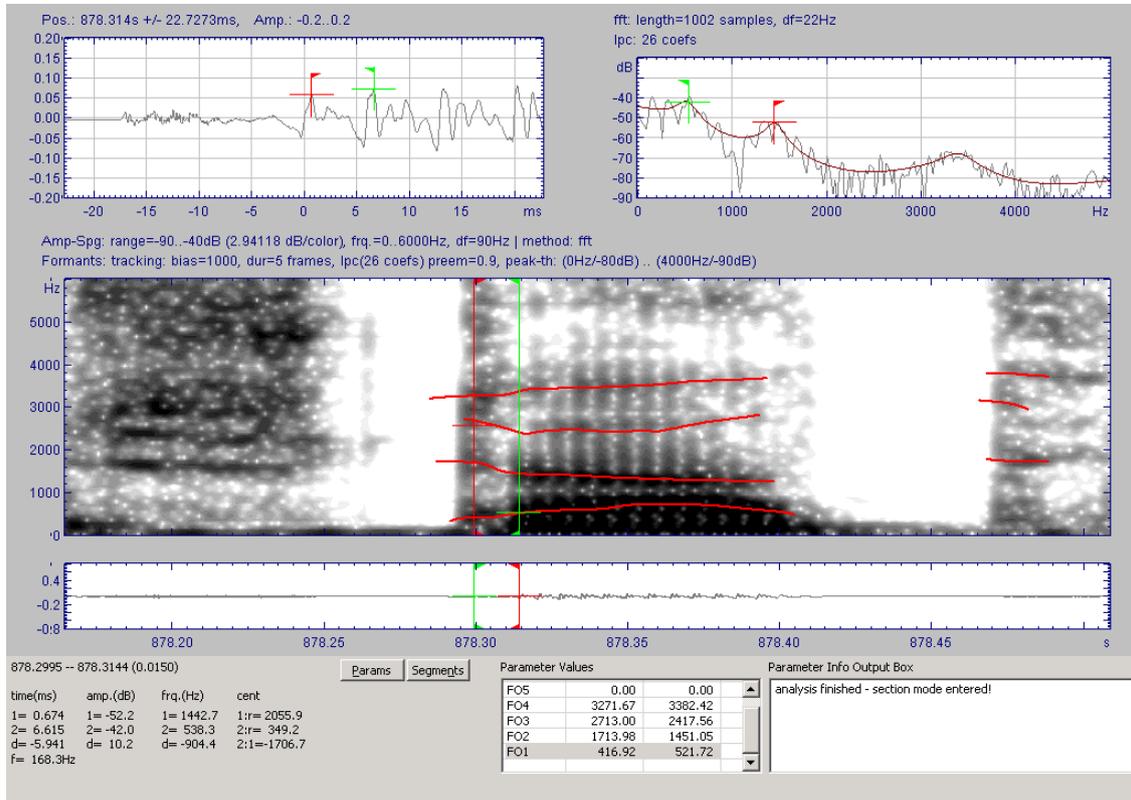


Figure 5.17: Spectrogram of the sequence /jtat/ from "Stadt" (city), speaker sp127, sentence reading task. Left cursor positioned during the burst of /t/, right cursor at vowel onset. Respective formant frequency values can be read from the panel below. Bottom panel: waveform window, next panel from bottom: spectrogram window, left upper panel: waveform zoom window, right upper panel: amplitude spectrum window.

It can be estimated from Figure 5.17, which shows the sequence [jɔ̌dat] "Stadt" (city), that the closure for the alveolar plosive is formed at about 2,5 cm posterior to the lip opening<sup>109</sup>. Therefore, as the tongue is withdrawn from this rather frontal position, a rapid and steep fall in F2 can be observed before vowel onset, followed by a monotonous fall from vowel onset till about vowel midpoint. Then the tongue body has to be fronted again for the following /t/. From modeling the sequence /da/, Manuel &

<sup>109</sup> From the spectral peak associated with F4.

Stevens (1995) could show that the fall of F2 is not caused by a change in the consonant constriction, i.e. by an opening of the closure; this increase of area at the point of closure location has hardly any effect on F2<sup>110</sup>. They conclude that the drop in F2 must be due to the tongue body moving back (Manuel & Stevens 1995: 439). However, the steep fall of F3 at consonant release until vowel onset and the rise of F3 from about vowel midpoint till vowel offset seems to be caused by the change in area at the point of consonant closure. Therefore, it seems to be possible to tease apart the components involved in a CV sequence where the plosive has an alveolar place of articulation and the vowel a constriction in the pharyngeal region: F2 seems to be associated with tongue body backing and fronting, F3 with change in constriction area.

Most of the work of tongue body backing is performed from the point of closure release until the onset of the vowel. However, at vowel onset, the tongue body is not yet in position. Therefore, some movement still has to be performed during the vowel.

As a closure in the velar region is formed, natural frequencies of the front and the back cavity come close together, very often resulting in convergence of F2 and F3. Such a convergence of F2 and F3 can be seen in the burst of the velar plosive /g/ in Figure 5.18, which shows the sequencs [gast] from "Gasthaus" (restaurant). After the burst, F2 and F3 drift apart again. F3 shows a slight rise and F2 drops slowly but continuously for about one third of the vowel, indicating tongue body retraction. This demonstrates that the tongue body does not take a back tongue body position in anticipation of the pharyngeal vowel (this would show up in a lower F2 spectral peak), but forms a closure in the velar region and retracts as the closure is released.

To date it is not clarified which factors are ultimately responsible for the forward positioning of the tongue in the production of velar plosives flanked by back vowels (called 'looping patterns'). The explanation put forward by Houde (1968, cited in Fuchs & Perrier 2005) that aerodynamics would cause the forward movement of the tongue

---

<sup>110</sup> This result of Manuel & Stevens (1995) is corroborated by the movement over time of F2 in Figure 5.17: F2 does not rise in preparation of the following alveolar plosive /t/.

has been partly refuted by Hoole et al. (1998) who observed looping patterns also in ingressive speech. Ohala (1983) attributed looping patterns to a strategy to maintain voicing. However, Mooshammer et al. (1995) found larger forward movements of the tongue for voiceless than for voiced velar stops. Löfqvist & Gracco (2002) postulated looping patterns to be planned in terms of cost minimization principles, whereas Perrier et al. (2003) and Fuchs & Perrier (2005) argue that looping patterns are due to biomechanical factors.

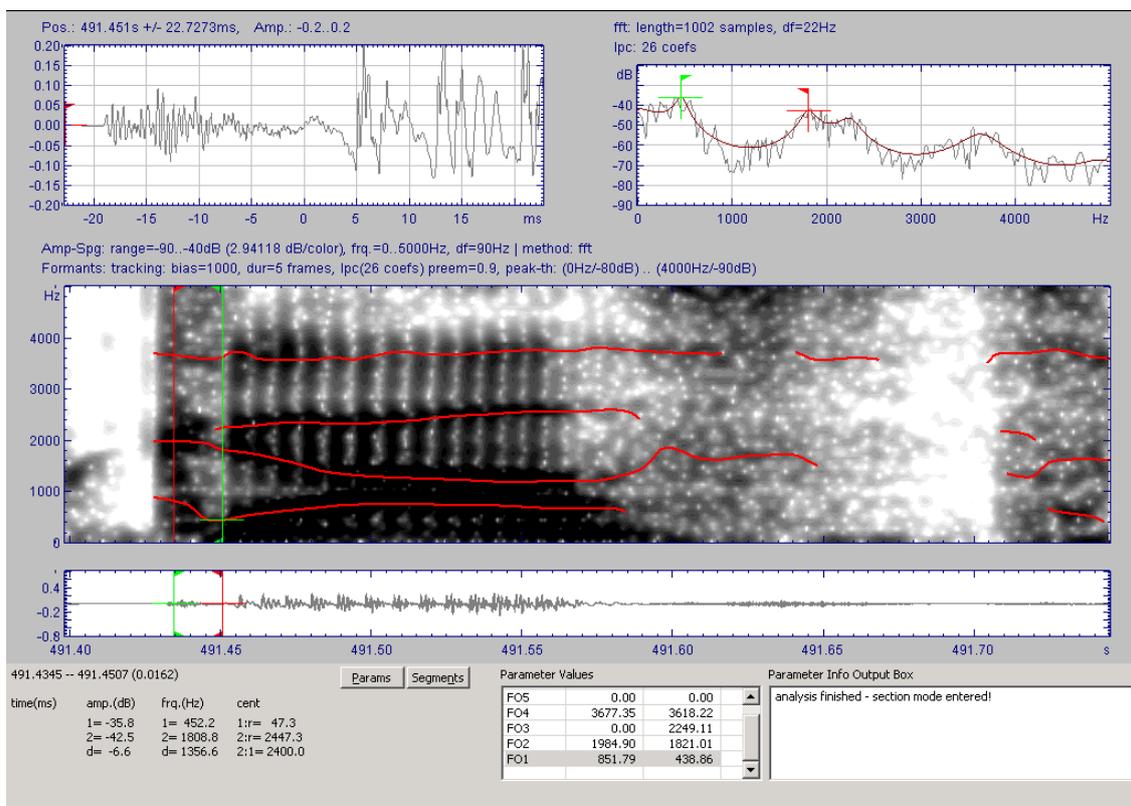


Figure 5.18: Spectrogram of the sequence /'gast/ from "Gast" (guest), speaker sp127, sentence reading task. Left cursor positioned during the burst of the velar plosive, right cursor at vowel onset of /a/. Respective formant frequency values can be read from the panel below. Bottom panel: waveform window, next panel from bottom: spectrogram window, left upper panel: waveform zoom window, right upper panel: amplitude spectrum window.

Therefore, due to the velar location, the spectral peak associated with F2 is higher than at an alveolar location, which in turn is again higher than for a bilabial location<sup>111</sup>.

<sup>111</sup> Resonance behind the closure for a tube closed at both ends.

Additionally, movements are slower in the case of a velar release. This adds to the fact that at vowel onset, F2 is higher when preceded by a velar plosive. Therefore, in the context of the vowel /a/, plosive place of articulation can be inferred from the F2 value at vowel onset (see Figure 5.19).

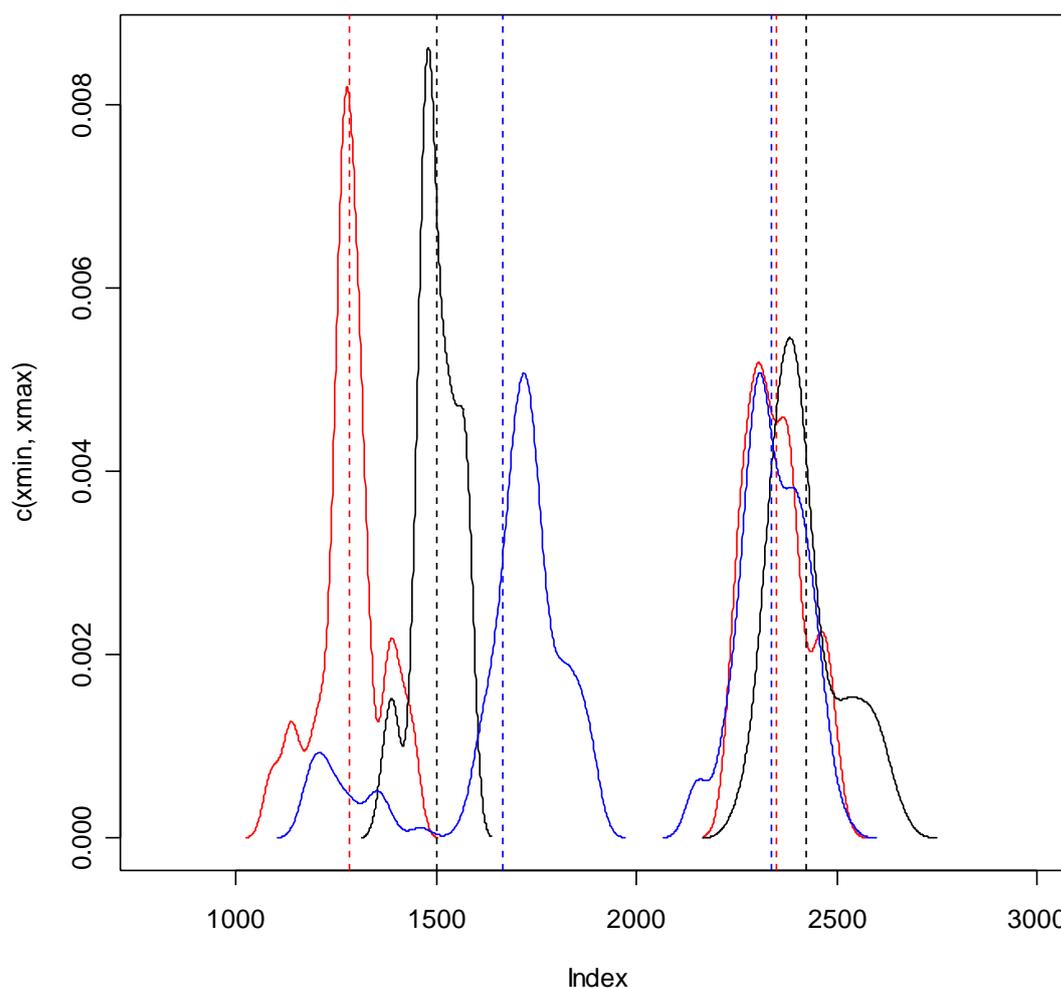


Figure 5.19: Density plot of the mean values of F2 and F3 of the first five frames calculated from vowel onset of all /a/ vowels in stressed positions, sentence reading task, speaker sp082. Red lines: C = bilabial, black lines: C = alveolar, blue lines: C = velar.

It can be seen from Figure 5.19 that F2 at vowel onset (the first five frames) is highest when preceded by a velar plosive, intermediate when preceded by an alveolar plosive

and lowest when preceded by a bilabial plosive. This pattern can be observed for all speakers, Table 5.5 shows that the differences in F2 at vowel onset are statistically significant.

mean F2	/ba, pa/	/da, ta/	/ga, ka/	F	p
sp180 <sup>112</sup>	1399	1724	1767	2;551:392,15	0.00
sp129	1184	1589	1773	2;606: 553,06	0.00
sp082	1292	1548	1674	2;531: 286,48	0.00
sp012	1176	1393	1475	2;432: 187,87	0.00
sp126	1178	1391	1471	2;571: 300,74	0.00
sp127	1218	1372	1576	2;456: 255,26	0.00

Table 5.5.: Mean F2 values of the first five frames of the vowel /a/ calculated from vowel onset in bilabial, alveolar and velar context, all conditions and positions. F and p value from one-way ANOVA.

The results on /Ca/ sequences, where C is either a bilabial, an alveolar, or a velar plosive are summarized as follows:

- /a/ preceded by a bilabial plosive shows two patterns, either a fall of F2 from consonant release until vowel onset or a flat movement of F2 from release onwards. Therefore, tongue body positioning takes place either during the occlusion of the plosive or at release. In most cases, tongue body positioning is finished at vowel onset.
- Preceded by an alveolar plosive, the tongue body has to be withdrawn after release. Therefore, tongue body positioning is not yet finished at vowel onset and a monotonous drop in F2 can be observed which reaches substantially into the vowel.
- In the case of a velar closure, the tongue body does not take a back location, but forms the closure in the velar region, which causes a drop in F2 that reaches substantially into the vowel. Therefore, again, tongue body positioning is not yet accomplished at vowel onset. The values of F2 at vowel onset are, however, higher than in the alveolar context.

<sup>112</sup> For p180, alveolar and velar context do not differ.

- The differences of F2 at vowel onset are statistically significant and point to the place of articulation of the plosive.

### 5.3.3. The vowels /u, o/

Both the vowels /u/ and /o/, especially when articulated with lip protrusion, are articulated in an acoustically stable region. The nomograms calculated by Stevens (1999: 281) show that over a range of about 6 cm (2-8 cm from the glottis) F2 hardly changes, when lips are protruded. This is exactly the region where the constrictions are located for both /u/ and /o/. Since coarticulatory effects appear especially in movements of the second formant, and since there are hardly any differences in F2 between /o/ and /u/, it is deemed justifiable to treat /o/ and /u/ as one.

It has already been discussed that in sequences where C is an alveolar plosive, lip protrusion starts at release and takes some time till it is accomplished. Therefore, whether lip protrusion is finished at vowel onset depends largely on VOT. Lip protrusion could not be traced at the offset of the transconsonantal vowel.

It can be seen from Figure 5.20 that both in bilabial and velar context, F2 at vowel onset is substantially lower than in the alveolar context.

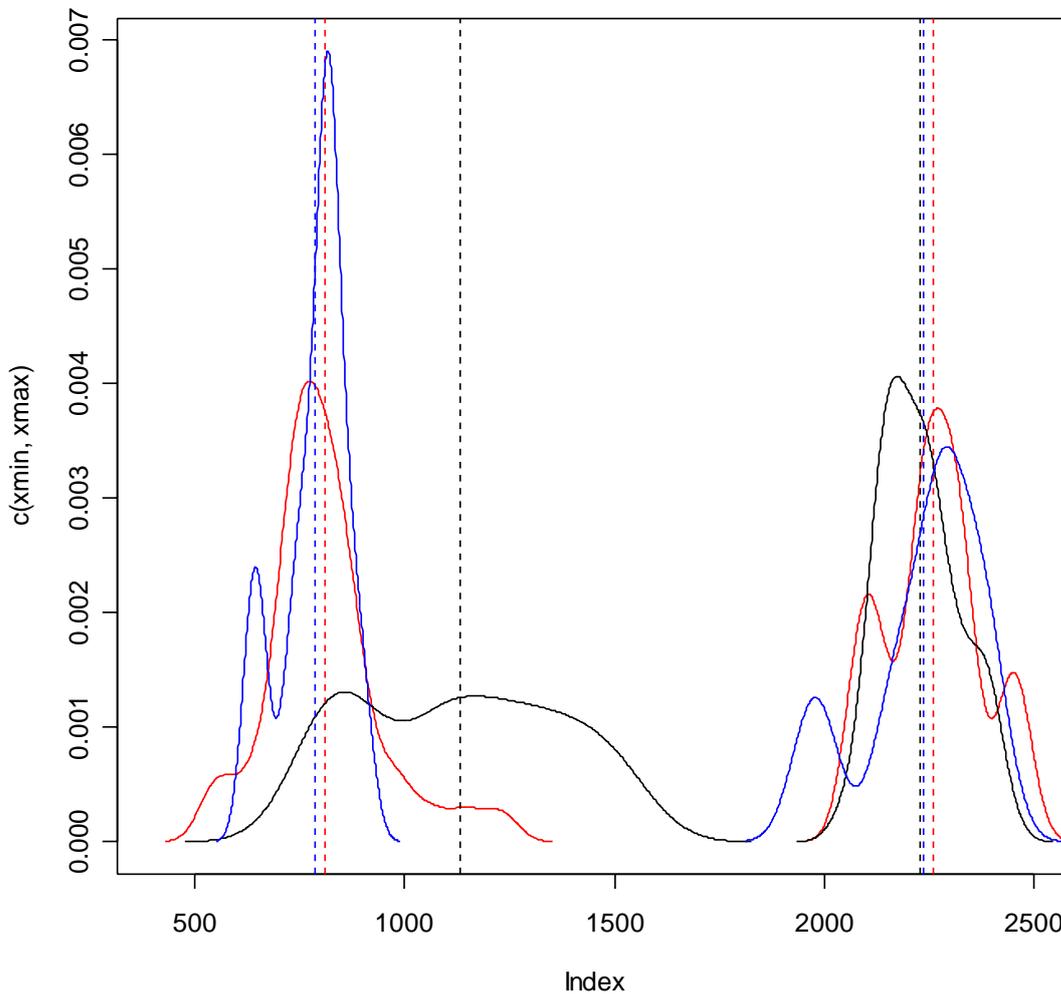


Figure 5.20: Density plot of the mean values of F2 and F3 of the first five frames calculated from vowel onset of all /u, o/ vowels in stressed position, speaker sp126, sentence reading task. Red lines: C = bilabial, black lines: C = alveolar, blue lines: C = velar.

However, the final values for F2 have not yet been accomplished at vowel onset in both cases. Figure 5.21 shows the sequence [b<sub>o</sub>:d<sub>ŋ</sub>] "Boden" (floor), cursors are positioned during the burst and at vowel onset.

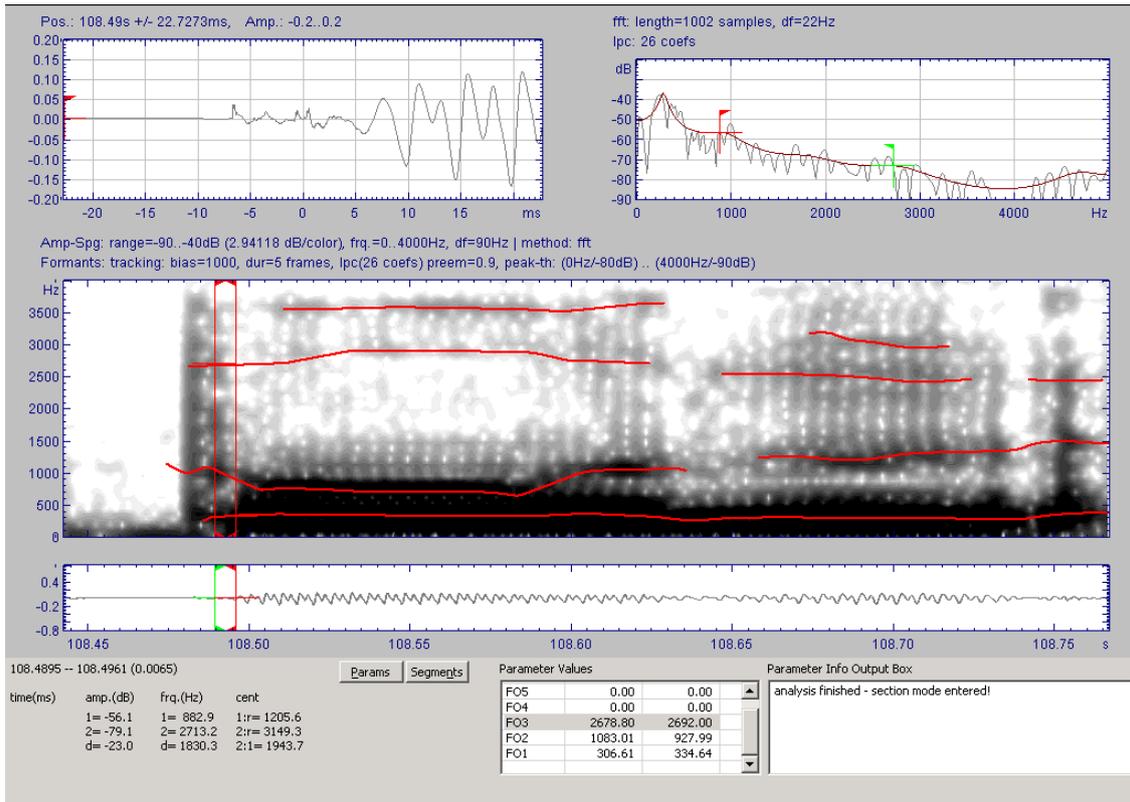


Figure 5.21: Spectrogram of the sequence [ˈbɔːdɪŋ] from "Boden" (floor), speaker sp180, sentence reading task. Left cursor positioned during the burst of the bilabial plosive, right cursor at vowel onset of /o/. Respective formant frequency values can be read from the panel below. Bottom panel: waveform window, next panel from bottom: spectrogram window, left upper panel: waveform zoom window, right upper panel: amplitude spectrum window.

It can be seen from Figure 5.21 that F2 still drops at vowel onset, indicating that lip protrusion is not yet accomplished at vowel onset, whereas the tongue body is most probably already in its position, as can be inferred from a comparison with F2 values for less protruded back vowels. In the case of /pɔːst/ "Post" (post office), taken from the same speaker, F2 has a value of 1062 Hz at vowel onset, this makes a difference of 135 Hz, which can be attributed mostly to lip protrusion. In Figure 5.21 F2, from vowel onset onwards falls another 150 Hz until the final position of F2 is reached. However, it can as well be the case that the spectral peak associated with F2 already has a sufficiently low value, so that it can be inferred that both tongue body backing and lip protrusion take place during the occlusion.

A similar pattern can be observed in where  $C_1$  is a velar plosive: the less protruded vowels /ɔ, ʊ/ have a F2 value of approximately 1050 Hz at vowel onset, with hardly any movement from release until vowel onset. In the case where a protruded vowel follows a velar plosive, it can be the case that F2 exposes either a sufficiently low value at release that it can be inferred that the tongue body is already in a back position for the velar closure (see Figure 5.22), or a drop in F2 can be observed at vowel onset.

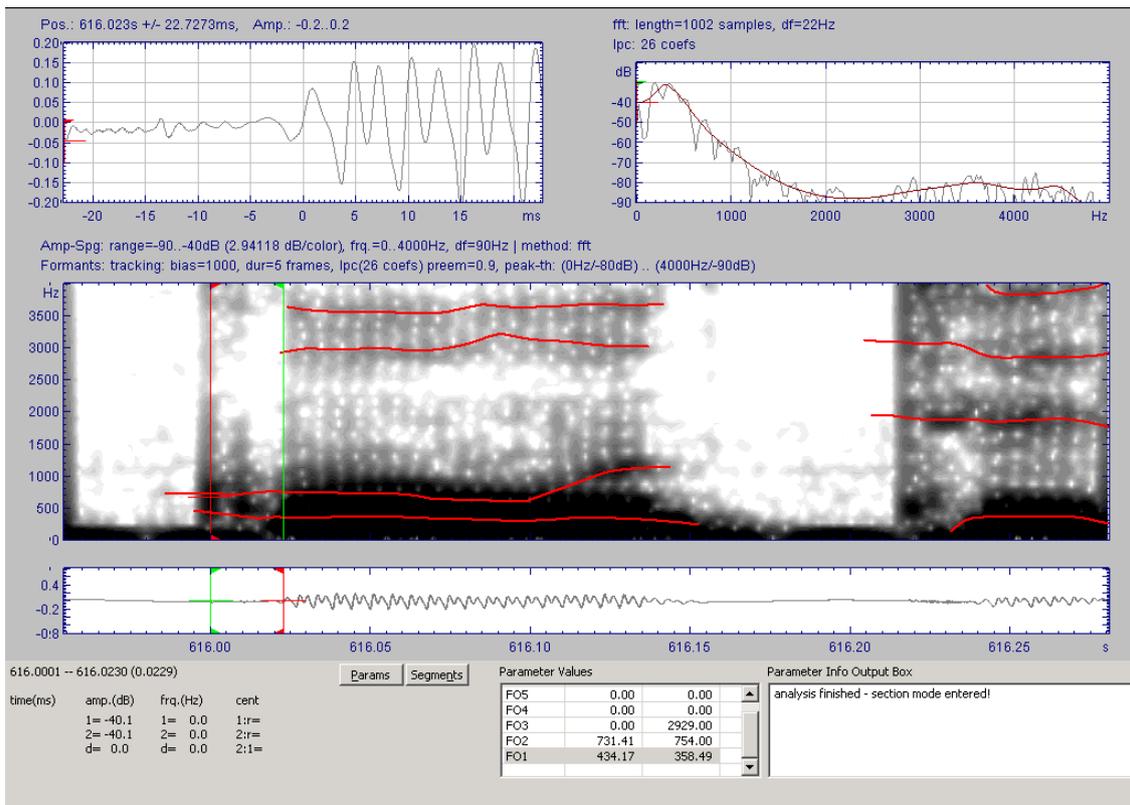


Figure 5.22: Spectrogram of the sequence /gotɛ/ from "Gote" (goth), speaker sp180, sentence reading task. Left cursor positioned during the first release of the velar plosive, right cursor at vowel onset of /o/. Respective formant frequency values can be read from the panel below. Bottom panel: waveform window, next panel from bottom: spectrogram window, left upper panel: waveform zoom window, right upper panel: amplitude spectrum window.

Figure 5.22 shows the sequence [gotɛ] "Gote" (goth); it can be seen that with a spectral peak at 730 Hz within the first burst, F2 is already sufficiently low. Consequently, no movement of F2 can be observed either during release or during the first two thirds of the vowel. Only when the tongue body is fronted to form an alveolar closure for /t/,

does F2 rise. However, a comparison of F2 at vowel onset (the first frame) with frame 20 (approximately vowel midpoint, depending on the total length of the vowel) of the vowel shows that F2 still drops after vowel onset (see Table 5.6).

F2 /u, o/	frame 1	frame 20	t	p
sp180	1063	777	3.28	0.00
sp129	786	626	2.88	0.00
sp082	1106	763	4.94	0.00
sp012	934	620	3.53	0.00
sp126	1040	693	6.22	0.00
sp127	1097	649	8.67	0.00

Table 5.6: Frame 1 and frame 20 F2 values for /u, o/ preceded by a bilabial, an alveolar or a velar plosive, sentence reading task. t value and p from one-tailed t-Test.

The results on /Cu/ - and /Co/ sequences, where C is either a bilabial, an alveolar, or a velar plosive are summarized as follows:

- F2 values are lower, when the back, protruded vowel is preceded by a bilabial or velar plosive.
- Two patterns can be observed: tongue body and lip protrusion have either been accomplished at release, or the tongue body is positioned after release. In the latter case, lip protrusion is pulled into the vowel.
- Lip protrusion causes F2 to drop approximately about 300 Hz.
- When preceded by an alveolar plosive, F2 is higher, indicating that tongue body retraction has not yet been accomplished at vowel onset. Therefore, both lip protrusion and tongue body positioning is pulled into the vowel.

## 5.4. Conclusion

It has been argued in the previous chapter that the way from phoneme to phonetic output is mediated by phonological processes, phonetic processes and coarticulation. Processes, in any case, are planned and processed before coarticulation is performed. This restricts the notion of coarticulation to the very last step on the way from one

phoneme, formed by either phonological or phonetic processes, to the next. The processes are language specific and, in some cases, serve to smooth away coarticulation – as could be seen in the case of palatalization. The sequence of a plosive and a following palatal vowel would cause a very steep and rapid transition, this transition is smoothed away by enlarging the contact area of the plosive preceding a palatal vowel. Whilst in Austrian German, this process only affects the neighbouring segment, in another language this process might be carried over into the transconsonantal vowel as well. In Russian, on the other hand, unpalatalised consonants are not palatalised when preceding a palatal vowel, it is – on the contrary – the palatal vowel that changes its quality in unpalatalised context (see Moosmüller 2007a for a discussion). Whether a given process is applied or not and the specific timing of this process in the case of application is responsible, among other things, for what is commonly termed as the "accent" of a given language or variety.

## 6. Vowel and Vowel Variability

The discovery of low-level variability as put forward by Menzerath & de Lacerda (1933) and Menzerath (1935) caused some consternation among the then phoneticians, “since it had often been assumed that the same sound would be articulated in the same way irrespective of its context” (Löfqvist 1997: 407). Since then,

“the overarching challenge was to explain the strong context-dependence and variability of acoustic phonetic patterns. Major handbooks (Hardcastle & Laver 1997) and review chapters (Farnetani 1997, Farnetani & Recasens 1999, Kent, Adams & Turner 1996, Löfqvist 1997) converge in identifying coarticulation as a major contributor to the mismatch between the dynamic and the linguistic perspectives of speech.” (Lindblom 2004: B-86)

The mismatch, however, lies much deeper than in the observation of variability in speech<sup>113</sup>. It lies in the conception of what is part of a grammar and what is not, and it lies in the unfortunate coincidence of the emergence of phonological theories which ascribed phonetic observations to performance and thus considered them irrelevant for linguistic investigation<sup>114</sup>. This incompatibility of phonology within generative frameworks and phonetics led to the strict rejection of phonology in the Eighties (Löfqvist 1986, Fant 1986):

“The supposedly happy marriage between phonology and phonetics has its inherent shortcomings and some of us like Peter Ladefoged might argue for a respectful divorce. [...] To me, phonetics is the stable partner of the marriage, while phonology is promiscuous in its experimenting with widely different frameworks and choice of features for describing one and the same inherent phenomenon.” (Fant 1986: 481)

Up to the present, the situation has not much changed. Mainstream phonological theories like, e.g., Optimality Theory, which incorporate phonetic knowledge and give elegant descriptions, are, however, of little help when addressing the main questions in phonetic theory (see Hurch 1998). While non-mainstream theories are seldom absorbed

---

<sup>113</sup> It strikes as somehow strange to blame an observation for the incompatibility between theories or theorists.

<sup>114</sup> To name just a few, see Dressler 1984, 1985, 1996, Hurch & Rhodes 1996, Donegan 2002 for a profound critique on phonological theories within generative frameworks.

into phonetics, some of them, such as Natural Phonology, incorporate phonetic results. Keating (1990) made a first step forward by incorporating phonetics into grammar. She advocates a strict separation of phonology and phonetics, in order to prevent confusion of the levels:

“Much of the coarticulation literature is confusing on this issue of levels, in that phenomena that are clearly phonetic are often given (unsatisfactory) phonological treatments. [...] The point is to determine the nature of each case.” (Keating 1990: 453)

Despite this strict introductory separation, it is not quite clear what the phonological and the phonetic components within her window model should be. The representation of the phoneme is the window which contains all information necessary to produce a given output.

“Context, not idealized isolation, is the natural state of segments,...” (Keating 1990: 461)

The window is empirically defined by the minimum and maximum values of a certain parameter (e.g. velum height, jaw position, or formant frequencies), the actual output is the path through successive windows, which is defined by the context (e.g. narrow windows allow less variability, wide windows allow more). Keating does not speak of rules, but rather it is the path which the articulators have to find through the successive narrow or wide windows that is responsible for the actual output.

In many ways, Keating’s model is convincing since it can account for much of the variability observed in actual speech production. The fact that the whole bulk of variability is stored in the mental representation is no counterargument, in the same way as economy of storage is no argument either. Nevertheless, some questions remain open, e.g. how do the different windows defining one specific segment interact and result in what has been termed motor equivalence (Perkell 1980, Abbs 1986)? And, finally, given a certain observed output, how is it to be decided whether it belongs to phonology or to phonetics?

It has already been argued in the previous chapter that much of what has been called “coarticulation” is in fact governed by processes (e.g. anticipatory velum

lowering or lip protrusion), which are language, variety, or even speaker specific, and, consequently, planned. Much of the debate within phonetic theory resides on the question of “what is controlled and what is a product of execution” (Lindblom 2004: B-86), i.e. what is not controlled. That such a question is posed at all has, in my opinion, its roots in the fact, that, all of a sudden, phoneticians were confronted with a huge amount of variability (e.g. vowels overlapping in formant frequencies) which they were unable to explain. Stevens & House (1963) and Lindblom (1963) explained the observed variability in their data within the framework of articulatory or acoustic undershoot. The notion of undershoot, on the other hand, implies the notion of a target that is undershot. The notion of the target, since it is usually conceived of as the intention of the speaker, is often intermingled with the notion of the phoneme. This inevitably causes an intermixture of levels.

## **6.1. Target, invariance, and target undershoot**

Stevens & House (1963) showed that consonant environment affects the realization of vowels. They interpreted their findings in terms of a production undershoot model. In the same vein, Lindblom (1963), in his work on vowel reduction, conceptualizes reduction in terms of undershoot. Additionally, he gives a concise definition of the target of a vowel:

“A target was found to be independent of consonantal context and duration and can thus be looked upon as an invariant attribute of the vowel. Although a phoneme can be realized in a more or less reduced fashion, the talker’s “intention” that underlies the pronunciation of the vowel is always the same, independent of contextual circumstances. A vowel target appears to represent some physiological invariance.” (Lindblom 1963: 1778).

According to this definition, the target is identical with the phoneme. Moreover, the phoneme is defined as a talker’s intention, the talker’s intention is invariant and it is assumed that the target corresponds more or less to a pronunciation under ideal conditions. In conditions deviating from this ideal state, the articulators fail to reach the intended target:

“Articulators respond to control signals not in a stepwise fashion but smoothly and fairly slowly, owing to intrinsic physiological constraints. Since the speed of articulatory movement is thus limited, the extent to which articulators reach their target positions depends on the relative timing of the excitation signal.” (Lindblom 1963: 1778)

This failure to reach the intended target position came to be known as “undershoot” (Stevens & House 1963, Lindblom 1963). To model the speech variability within the framework of articulatory/acoustic undershoot and corresponding perceptual overshoot (Lindblom & Studdert-Kennedy 1967, Nearey 1989, see van Son 1993 for an overview) is convincing, satisfactorily explains variability, and is, today, a well-established concept in phonetics. Yet, the question

“whether and to what extent the human speech perception could be able to recover intentions in motor tasks that are not achieved” (Perrier 2005: 128)

is still unsolved. Moreover, given the fact that the actual output does, in most cases, not correspond to the intended target, especially in weak prosodic positions, speech production can be considered a consecutive succession of failures. Speaking, within this framework, would be an utterly frustrating activity. Apart from that, languages display very clear sequential constraints, in the way that unpronounceable sequences are not allowed. Therefore, among the languages described so far, no sequence of phonemes such as, e.g., \*/fpdk/, can be found. It has to be questioned why a language should allow a sequence for which the articulators constantly and systematically fail to reach the intended target.

One crucial aspect in Lindblom’s argumentation for the undershoot concept is duration:

“As a vowel becomes shorter, there is less and less time for the articulators to complete their “on-“ and “off-glide” movements within the CVC syllable. Provided that the neural events corresponding to the phonemes actually stay invariant, the speech organs fail, as a result of the physiological limitations, to reach the positions that they assume when the vowel is pronounced under ideal steady-state conditions.” (Lindblom 1963: 1779)

Subsequent studies showed that vowel quality does not change as a function of vowel duration (Gay 1978, Nord 1975, 1986, Engstrand 1988, van Son & Pols 1990, 1992, van

Son 1993<sup>115</sup>). Gay (1978) showed that when the vowel is shortened, it is predominantly the onset that is changed, whereas the vowel midpoint stays invariant. For the vowel /i/, Gay states:

“For all subjects, the onset frequency of the second formant transition is higher for the fast rate condition, while the F2 midpoint frequencies and F2 rates of change remain essentially unaffected across the two rates.” (Gay 1978: 226)

However, the acoustic invariance is not accompanied by articulatory invariance:

“The present acoustic data are also inconsistent with earlier EMG data (Gay et al., 1974, Gay and Ushijama, 1975) that showed a change in the level of muscle activity for vowels in response to a change in speaking rate. The EMG data showed that the activity levels of the genioglossus muscle for the vowel /i/ decreased with an increase of speaking rate”. (Gay 1978: 228).

Compensatory articulatory strategies for guaranteeing a desired acoustic output are very well documented. Perturbation experiments (Savariaux et al. 1995, 1999, Hoole 1987, Jones & Munhall 2003) show that subjects try to compensate the perturbation in order to ensure an improvement of the acoustic output<sup>116</sup>. Men and women use different strategies in the production of back rounded vowels, which, in both cases, guarantee a sufficiently low F2 (Fant 2004). Furthermore, speakers adjust their articulation according to individual vocal tract shapes. Perkell (1997) could show that tongue height in the production of the vowels /i, ɪ, ε/ differed according to the individual shapes of the palate.

“The speaker with the shallowest palate vault (1) uses the smallest adjustments of tongue height to create the area function differences required for the vowels, and the speaker with one of the steepest vaults (3) uses the largest adjustments.” (Perkell 1997: 350)

Similar results were obtained by Pouplier et al. (2004) in their study on the tense – lax distinction in German:

“There is considerable inter- and intrasubject variability as to the palatal distance within a tense and lax pair.” (Pouplier et al. 2004: 24)

---

<sup>115</sup> van Son (1993) concluded that target-undershoot is most probably planned.

<sup>116</sup> Löfqvist (1997) argues that the fact that speakers needed several trials to obtain the desired acoustic output, speaks against acoustically-based targets, whereas for Perkell (1997) several trials are no counterargument against acoustically-based targets and are compatible with this concept, because, in bite-block experiments, subjects acquired an alternative set of motor command-to-acoustic mappings and retained this set for later recall (1997: 364)

A subsequent study (Brunner et al. 2005) showed that subjects with a flat palate are allowed less articulatory variability because slight displacements would result in great changes in the acoustic output, whereas subjects with dome shaped palates are allowed more articulatory variability, which they can employ or not.

These results strongly suggest that “the objective of speech articulation is to produce an acoustic signal with properties that will enable the listener to understand what is said” (Perkell 1997: 363). Within the Acoustic Invariance Theory, some of these properties have to be invariant. Blumstein & Stevens (1979), who showed that the stop place of articulation can be arranged according to spectral properties of the burst (diffuse-raising, diffuse-falling and compact), conclude:

“In particular, it has been shown that the speaker provides the listener with invariant acoustic cues, cues which can be directly derived from the speech signal itself. Thus, the interface between the perceptual and production systems resides in the acoustic signal where the properties of speech can be uniquely and invariantly specified.” (Blumstein & Stevens 1979: 1015)

Despite these initially promising results with respect to the stop place of articulation (see also Blumstein & Stevens 1980, Stevens & Blumstein 1978, Lahiri et al. 1984, Blumstein 1986), the search for an invariant acoustic property was unsuccessful. Löfqvist (1986) argued for a separation of phonetics and phonology in order to get away from a concept that sees invariance in terms of static entities. In a similar way, Fant (1986) doubts “the absolute invariance of feature correlates irrespective of context” (1986: 486) and suggests a context-dependent analysis instead:

“In my view, human speech perception relies on gestalt decoding rather than on isolated short-time spectral patterns or templates. [...] The auditory system probably makes efficient use of the entire evidence available. Why should we limit our descriptive work to less precise specifications or to a diluted specification which can operate in all contexts?” (Fant 1986: 487f)

Yet, results on the role of context in vowel perception are not unanimous. The theory of the dynamic specification of vowels (Strange 1998, Strange & Bohn 1998, Jenkins et al. 1999) states that vowels in continuous speech are better identified when context is available, whereas the results of Nearey & Assmann (1986) and Andruski & Nearey

(1992) showed that listeners rely more on vowel-inherent factors and that coarticulatory cues play a minor role in the perception of vowels. Hillenbrand & Nearey (1999) conclude that

“spectral change patterns play a secondary but quite important role in the recognition of vowel quality. [...] However, a simple observation that should not be lost in this discussion of spectral change is that the single-slice spectral measurements reported in studies such as Peterson and Barney (1952) capture most of the information that is needed to represent vowel quality. In the present study, F0, duration, and steady-state formant measurements were sufficient to signal the intended vowel for roughly three-fourths of the utterances, with nearly all of the misidentifications involving adjacent vowel categories.” (Hillenbrand & Nearey 1999: 3521)

Pitermann (2000) comes to the conclusion that static information is sufficient for identifying the vowels. Carré & Divenyi (2000) stress the importance of dynamic changes.

It is already a challenge to look for invariant acoustic properties for stop consonants. Modifications to the original templates proposed by Blumstein & Stevens (1979) have been made in order to extend the concept to languages which differentiate further or places of articulation other than the bilabial, alveolar or velar (Lahiri et al. 1984, Blumstein 1986). This task seems even more unsolvable for vowels. As is well known, the acoustic properties of vowels change in dependence on linguistic and extralinguistic factors, languages or language variants. Thus, the acoustic properties of the vowel /i/ are not the same in English and Standard Austrian German; in English, /i/ is characterized by a spectral dominance of F2 and F3, in Standard Austrian German by a spectral dominance of F3 and F4. Therefore, the claim of the theory of acoustic invariance,

“that a particular phonetic dimension should be realized by the same invariant property across all languages” (Lahiri et al. 1984: 391),

cannot be upheld for vowels. Confronted with this lack of (articulatory and acoustic) invariance, it has to be asked, consequently, what makes an /i/ an /i/ (see also the discussion in Donegan 2002).

## 6.2. Undershoot vs. Processes

A frequently cited example for undershoot is the fronting of back rounded vowels in the alveolar context in English. In their study on the effects of consonant environment on vowels, Stevens & House (1963) found that especially the F2 of the vowel /u/ was displaced by about 350 Hz in the alveolar context. The Stevens & House study was repeated by Hillenbrand et al. (2001). Their results were similar to Stevens & House's; as concerns the back rounded vowel, they found even larger F2 displacements:

“As with SH [Stevens & House], the largest effect by far is a raising of F2 for /u/ in the environment of alveolar consonants. At about 500 Hz for the men and nearly 600 Hz for the women (relative to null environments), this upward shift is even larger than the roughly 350 Hz effect reported by SH.” (Hillenbrand et al. 2001: 754)

This displacement of F2 is often interpreted within an undershoot model and allocated to physiological constraints (Chomsky & Halle 1968). Since the targets are a long way away from each other, the tongue cannot move sufficiently back in order to form a constriction in the velar region for the vowel /u/. Moon & Lindblom (1994), in analyzing /wVC/-sequences, concluded that

“from a biomechanical point of view, formant undershoot ought to be a function of “locus-target” distance, vowel duration and F<sub>2</sub> rate of change.” (Moon & Lindblom 1994: 53).

However, F2 displacement could not be observed for all languages. Flemming (2001) showed that French, German and Hindi display F2 undershoot to a lesser degree than English, with German displaying the least F2 displacement from the defined target<sup>117</sup>. It is further argued that languages with high front rounded vowels do not apply fronting in order to maintain the necessary contrasts. Within this line of argumentation, Oh (2002) showed that Chinese, French and German display considerably less F2 displacement for /u/ in alveolar context than English. However, even within these three languages, which all exhibit front rounded vowels, differences occurred:

---

<sup>117</sup> Target F2 was estimated by measuring F2 either in isolation or adjacent to [h] (Fleming 2001: 23).

“German was characterized by having the lowest  $F2_T$ <sup>118</sup> values and one of the smallest undershoot values of the back vowel in the coronal contexts. Regarding Chinese and French, while Chinese had a more front target for /u/ than French, the vowel /u/ of French was undershot more than that of Chinese in coronal contexts.” (Oh 2002: 244)

From a subsequent study on second language acquisition, Oh (2002) concludes that these language-specific differences have to be learned and interprets them within the framework of Keating’s model:

“These results are in accord with the Keating’s (1985) conception of the language-specific phonetic component of grammar which can be hypothesized that a speaker must learn all phonetic details that are specific to a target language.” (Oh 2002: 252)

With the results presented by Fleming (2001) and Oh (2002), a biomechanical explanation for the fronting of back rounded vowels can be excluded, therefore, no undershoot mechanism is at work. Oh (2002) interprets F2 displacement as a language-specific coarticulatory process which has to be learned. Fleming (2001), on the other hand, sees fronting as a parallel phonetic and phonological phenomenon and analyzes it within the framework of minimization of cost functions. As concerns the language-specific differences, he concludes:

“this variation can be analysed in terms of differences in constraint weights, e.g. English assigns IDENT(V) a low weight compared to German.” (Fleming 2001: 23)

In the sentence reading condition, Standard Austrian German shows no F2 displacement<sup>119</sup> of the vowel midpoint of the stressed back rounded vowels /u, ʊ, o/ following an alveolar plosive. However, statistically significant differences of F2 displacement caused by the preceding alveolar plosive can be found for the vowel /ɔ/. Broken down for speakers, every speaker exposes a higher F2 value at vowel midpoint when the preceding plosive is alveolar, and for all speakers, the differences are statistically significant ( $p < 0.05$ ). Table 6.1 gives the mean vowel midpoint values of the vowel /ɔ/ for each speaker:

---

<sup>118</sup>  $F2_T = F2$  target value

<sup>119</sup> It has to be noted that fronting is the most plausible explanation for a higher F2 of back rounded vowels in an alveolar context. However, less lip protrusion, a raised larynx, or an increase of constriction degree might, in the same way, result in a higher F2.

F2 /ɔ/	Sp012	Sp126	Sp127	Sp082	Sp129	Sp180
bilabial	760	870	786	986	788	1039
alveolar	874	1034	952	1136	924	1276
velar	793	846	734	1062	851	1014

Table 6.1: Mean F2 vowel midpoint values of the vowel /ɔ/ according to the place of articulation of the preceding plosive. Data broken down according to speakers, vowel in stressed position, sentence reading task.

The observed differences are independent of vowel duration, i.e. F2 at vowel midpoint does not raise as duration decreases. No statistically significant relationship between duration (expressed in NoP) and F2 at vowel midpoint was observable for any of the speakers (see Figure 6.1 and Table 6.2).

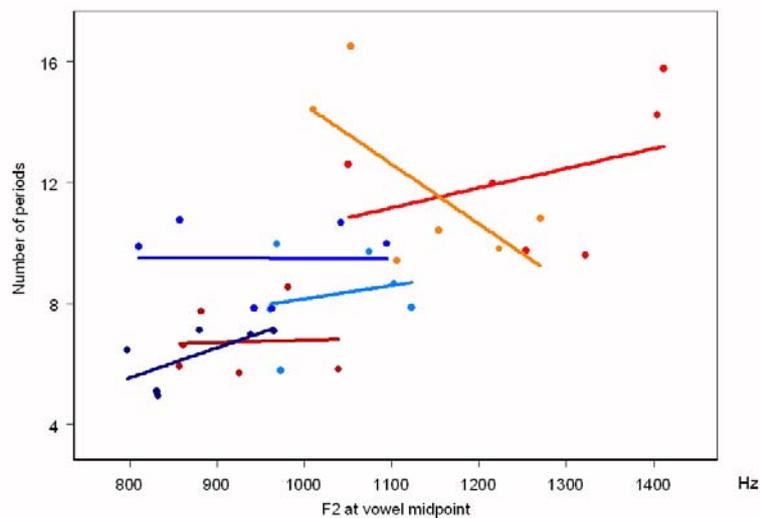


Figure 6.1: Regression lines and scatter plot of duration (number of periods) vs. F2 at vowel midpoint of CV – sequences, where C is an alveolar plosive and V a back, stressed vowel /ɔ/. Reddish lines: female speakers, bluish lines: male speakers, sentence reading task.

Table 6.2. gives the correlation coefficient  $r$  for all speakers:

Speaker	p082	p129	p180	p012	p126	p127
$r$	0.723	0.005	0.386	0.574	0.252	0.011

Table 6.2.:  $r$  for duration (number of periods)/F2 at vowel midpoint, broken for all speakers, stressed vowel /ɔ/ with preceding alveolar plosive, sentence reading task,  $p > 0.05$ .

In spontaneous speech, it was not possible to find sufficient items for all speakers. However, those speakers who uttered enough items within each of the three consonantal contexts, exposed no statistically significant differences for F2 at vowel midpoint. Moreover, F2 values at vowel midpoint dropped in spontaneous speech as compared to the sentence reading condition, adjusting – to those in a bilabial or velar context.

These results suggest investigating fronting in the most formal task, the reading of logatomes. Table 6.3 gives the results of one-way ANOVAs for the back vowels at vowel midpoint:

	Sp012				Sp180			
	/u/	/ʊ/	/o/	/ɔ/	/u/	/ʊ/	/o/	/ɔ/
F1	–	+	–	–	+	–	–	–
F2	–	+	–	+	–	+	–	+
F3	–	+	+	+	+	–	–	+

Table 6.3: Statistically significant differences ( $p \leq 0.05$ ) based on one-way ANOVAs of vowel midpoint formant frequencies of the back rounded vowels depending on the place of articulation of the preceding plosive (bilabial, alveolar, velar). Speaker sp012 and sp180, logatome reading task.

As concerns F2, both sp012 and sp180 expose statistically significant differences for the [–constricted] vowels /u/ and /ɔ/. Figure 6.2 shows the spectral change over time for all vowels /ɔ/ in the bilabial, alveolar, and velar context for p180.

It can be seen from Figure 6.2 that F2 values at vowel midpoint of almost all items in the alveolar context exceed the values in the bilabial and velar context, and the calculated F-value considerably exceeds the critical F-value ( $F_{(2,21)} = 33,32$ ,  $p = 0.00$ ). Again, no correlation could be observed between F2 at vowel midpoint and duration (expressed in NoP). The same can be observed for the vowel /u/ for both p012 and p180. Table 6.4 gives the mean F2 values measured at vowel midpoint for the back rounded vowels in bilabial, alveolar, and velar contexts.

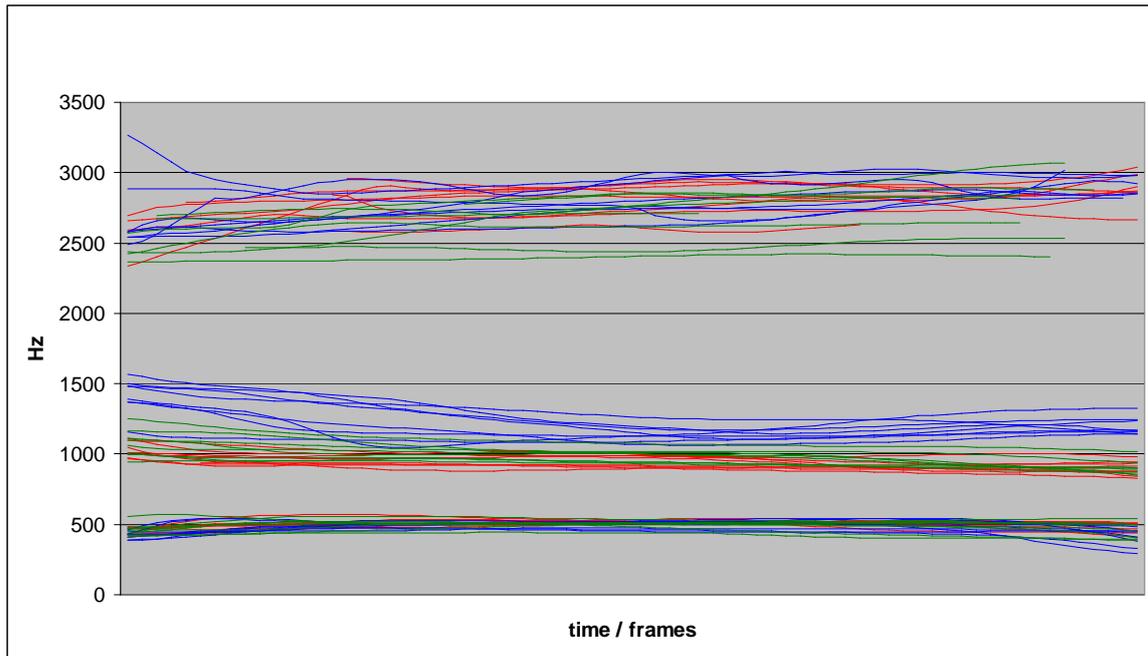


Figure 6.2: Linear time standardized spectral change over time (F1, F2, F3) of all vowels /ɔ/ in bilabial (red lines), alveolar (blue lines), and velar (green lines) contexts, speaker sp180, logatome reading task.

Sp012					Sp180			
F2	/u/	/ʊ/	/o/	/ɔ/	/u/	/ʊ/	/o/	/ɔ/
bilabial	562	<b>631</b>	561	<b>835</b>	749	<b>827</b>	767	<b>948</b>
alveolar	608	<b>828</b>	591	<b>927</b>	835	<b>1133</b>	773	<b>1153</b>
velar	540	<b>584</b>	544	<b>825</b>	781	<b>838</b>	688	<b>991</b>

Table 6.4: Mean F2 vowel midpoint values according to place of articulation of the preceding plosive. Speakers sp012 and sp180, logatome reading task. Vowels exposing statistically significant differences ( $p < 0.02$ ) for F2 are in bold.

It follows from quantal theory that lip protrusion and concomitant lowering of the larynx stabilizes the F2 of back vowels (Stevens 1972, 1989, 1999). Therefore, F2 is not to be affected by small changes in the constriction location. Moreover, a tighter constriction lowers F2 and raises F3 (Fant 2004) for the back vowels. Therefore, from F2, it cannot be decided whether fronting takes place in the production of /u/ and /o/ in the alveolar context. With less or no lip protrusion, however, F2 reacts to small changes in constriction location. This is the reason why F2 displacements can be observed for

the vowels /ʊ/ and /ɔ/, which expose less lip protrusion as compared to their counterparts /u/ and /o/.

F3, however, might be an indicator of fronting for the vowels /u/ and /o/; a higher F3 pointing to a more fronted articulation. As Table 6.3 demonstrates, statistically significant differences can be observed for F3 for the vowels /ʊ/ and /ɔ/ of speaker sp012 and for the vowel /u/ of speaker sp180. Table 6.5 gives the mean F3 values measured at vowel midpoint for the back rounded vowels in bilabial, alveolar, and velar contexts:

Sp012					Sp180			
F3	/u/	/ʊ/	/o/	/ɔ/	/u/	/ʊ/	/o/	/ɔ/
bilabial	2374	2485	<b>2637</b>	<b>2195</b>	<i>2783</i>	2449	2669	2809
alveolar	2550	2478	<b>2778</b>	<b>2279</b>	<i>2292</i>	2648	2556	2814
velar	2511	2313	<b>2536</b>	<b>1914</b>	<i>2651</i>	2577	2885	2672

Table 6.5: Mean F3 vowel midpoint values according to place of articulation of the preceding plosive. Speakers sp012 and sp180, logatome reading task. Vowels exposing statistically significant differences ( $p < 0.01$ ) for F3 are in bold, vowels which show no higher values for F3 in alveolar context, are additionally in italics.

Except for the vowels /o/ and /ɔ/ of speaker sp012, which might indicate some fronting, the values of F3 are not conclusive; i.e. they are not higher in the alveolar than in the bilabial or velar context. It can be concluded that in the most formal condition, F2 displacement can be observed for the vowels /ʊ/ and /ɔ/. The highest amount of F2 displacement can, therefore, be observed in the most controlled task, the reading of logatomes, followed by the sentence reading task, where displacement of F2 can only be found in the vowel /ɔ/. In spontaneous speech, the differences no longer exist. What can be observed is a hierarchical diminution of differences in dependence on the speech situation or speaking task. This strongly suggests that F2 displacement is neither a matter of undershoot nor a coarticulatory phonetic detail, but rather a process which uses a phonetic circumstance to maximize contrast in most formal speech situations. Higher articulatory variability of citation forms as compared to, e.g., read sentences, has

already been observed elsewhere (e.g. Abbs 1986). Vaissière (1986) attributes this lack of invariance partly to non-linguistic factors and partly to “context-dependent articulatory “weights” assigned by the speaker to different parts of the syllable.” (Vaissière 1986: 222). The results of Scarborough (2004) are in the same vein: “hard”<sup>120</sup> words expose a higher degree of coarticulation than “easy” words. Moreover, she proved that the presence of coarticulation facilitates perception, and that, consequently, listeners make use of coarticulatory information. These findings together with the observed language-specific differences rule out biomechanical factors and suggest that processes are at work. Moreover, within the figure-ground preference of Natural Phonology (Dressler 1984, 1986, 1996), the analysed phenomenon can be characterized as a foregrounding process.

### 6.3. Phoneme, allophone, target, and processes

The examples discussed in 6.2 show that the target is variant and has to be conceptualized differently from the phoneme which is invariant (see 1.2). The phoneme is the mental representation of a sound and can be modified by processes. The final output of the phoneme is usually hit by speakers with no speaking or hearing deficits and under normal speaking conditions. In this way, the final output is an invariant target, cf. Keating (1990):

“... any single given context reduces, not introduces, variability in a segment” (Keating 1990: 461).

However, since Keating does not assume processes or rules, she allows “arbitrary variation” (1990: 467). The example given for arbitrary variation is Russian, “with extensive vowel allophony and reduction” (1990: 467), modelled by defining wide

---

<sup>120</sup> Hard words are words, which, with respect to their intelligibility, expose a low frequency relative to the sum of the frequencies of all their neighbours and are therefore obscured by their neighbours, whereas easy words have a high frequency relative to the sum of the frequencies of all their neighbours and – consequently – stand out against their neighbours.

windows for Russian vowels. Russian vowel allophony already highly challenged, if not defeated by Öhmann's (1966) vowel-to-vowel coarticulatory model. Yet, conceptualizing as residual "arbitrariness" that which cannot be explained in the same way as for Germanic languages, should be avoided in linguistic theories.

Russian vowel allophony, however, is highly systematic:

"We agree that the variability of Russian vowels is tremendous but at the same time is of systematic character that must and can be captured by systematic and scrupulous study of all the aspects of vowel functioning in speech under strict control of relevant variables." (Kouznetsov 2001: 439)

It has been shown in the previous chapter that the vowel /i/, which is usually attributed a high coarticulatory resistance (Recasens 1999, Fowler & Brancazio 2000), palatalizes the previous plosive in Standard Austrian German. Therefore, the vowel /i/ is not affected by the preceding plosive.

In Russian, with its palatalised and non-palatalised consonants, the situation is inverse. The palatalised – non-palatalised opposition has to be preserved, irrespective of the vocalic environment. If, in the same way as in Standard Austrian German, /i/ would palatalize the non-palatalised consonant, a phonemic contrast would be lost. Therefore, it is the consonant that changes the vowel. In a palatalised context, the quality of /i/ is of course preserved. But,

"/i/ following a *non*palatalized consonant is repelled from the high-front position in the vowel quadrilateral and acquires a [i]-like (or even [ɪ]-like) onglide. It is usually said that Russian /i/ has the allophone [i] following a nonpalatalized consonant, [...] F2 of /i/ tends to follow a rather sigmoid course following a nonpalatalized consonant, having only a slight slope at the release of the consonant, then rising fairly steeply, then, finally, leveling off again toward its target." (Howie 2001: 18)

The difference between Standard Austrian German and Russian<sup>121</sup> lies in the language specific phonologies with resulting different processes. A process which changes the quality of the vowel /i/ preceded by a non-palatal consonant (Russian) is as "natural" as a process changing the quality of the consonant preceding the vowel /i/ (Standard Austrian German). The phonology of a language decides which processes have to be

---

<sup>121</sup> Kouznetsov (2001) analyzes speakers from Moscow, Howie (2001) does not further specify the variety he analyses and models.

suppressed and which are allowed, in order to maintain the relevant oppositions<sup>122</sup> (see also Donegan 2002). Therefore, it is not a segment which does or does not exert a certain influence on contiguous segments, but the phonology of a language which decides on the status of the phonemes and, hence, whether they are allowed to exert an influence or not. From this follows, logically, that a certain output is not left to some uncontrolled biomechanical circumstances, but planned and exerted via processes.

“Speakers do not simply line up a sequence of phonemic targets and allow the articulators to get from one to another as best they can; instead, the activity of articulation is centrally planned, so that features spread (or gestures overlap) in regular ways. This planning differs from language to language, [...]” (Donegan 2002: 69)

Therefore, the whole string from the initial planning of an utterance up to its final output is – under the above-mentioned conditions – planned and under the control of the speaker. In this way, phonetics is part of the grammar.

## 6.4. Missed targets

### 6.2.1 Speech errors

Cases where the target is not hit, exist of course. However, under “normal” conditions, these cases are rare and can be subsumed under what is usually called speech errors. Figure 6.3 gives an example of such an error, where the target vowel quality is not reached.

In this example, the speaker intends to read the utterance „Heinz regelt das für dich, er wird ein Lokal finden“ (Heinz will settle this for you, he will find a good place). For some reason, the speaker does not succeed in producing the right target for the vowel /ɪ/. Instead, he realizes the vowel /ɛ/. He stops immediately after he had realized that he had produced the wrong vowel and, after a short break, starts with a second trial for the word “dich” (you). A comparison of the formant frequencies in the vowel /ɛ/

---

<sup>122</sup> The differences between Russian and Standard Austrian German vividly show that there are no “default values” (Fowler 1990) which need not be modelled via processes.

(“de-“) and the vowel /ɪ/ (“dich”) reveals the undershoot. Therefore, a speaker immediately notices an articulatory and/or an acoustic undershoot and corrects this failure.

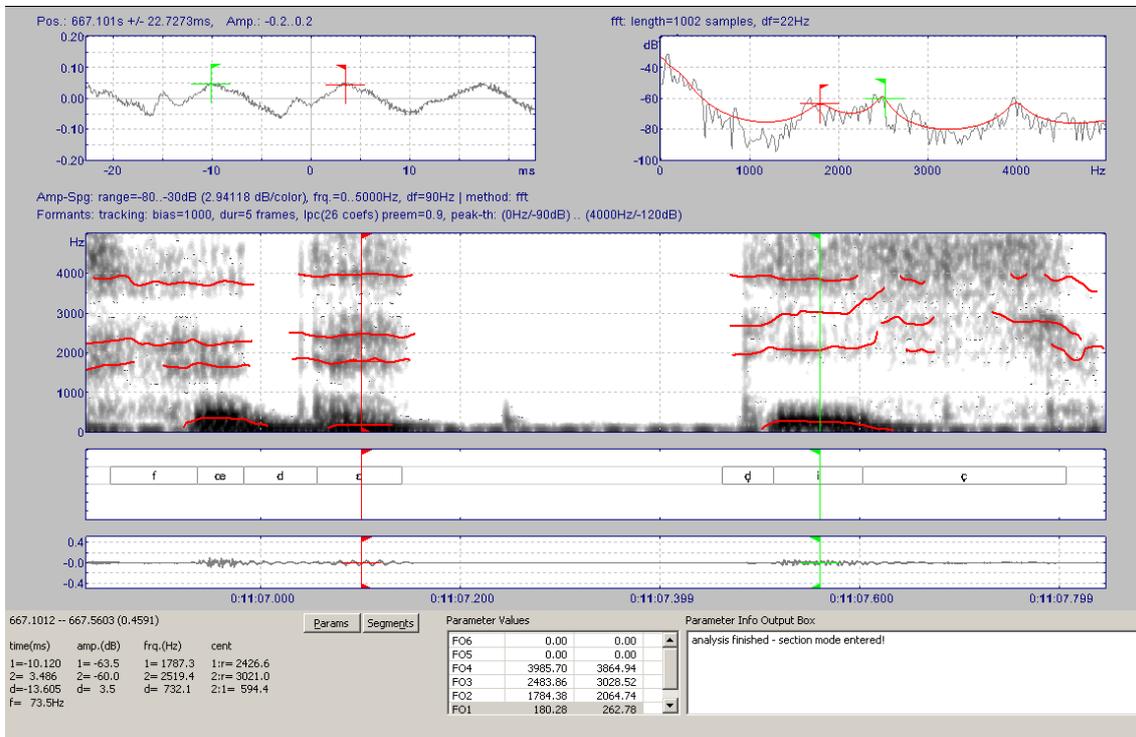


Figure 6.3: Spectrogram of the sequence /fyr dɪç/ “für dich” (for you), speaker sp012. Cursors positioned at vowel midpoint, respective values of formant frequencies can be read from the panel below. Bottom panel: waveform window, next panel from bottom: phonetic transcription<sup>123</sup>, 3<sup>rd</sup> panel from bottom: spectrogram window, left upper panel: waveform zoom window, right upper panel: amplitude spectrum window.

## 6.4.2 Articulatory avoidance

In speaker identification, reference voices of a speaker have to be recorded, in order to be able to compare two speakers. In this situation, some speakers might become very inventive in how to disguise their natural speaking mode. One strategy is to avoid articulation. This avoidance of articulation can be discerned from casual speech insofar,

<sup>123</sup> After the undershoot, the speaker makes a break. Therefore, the start of the closure for the subsequent plosive cannot be made out. The left boundary of the /d/ in the transcription line is, therefore, not meant to mark the start of the closure phase.

as the output in the first case is inconsistent. The reason for this inconsistency lies in the fact that the speaker deliberately tries to change the plan: the output target is to depart from a neutral vocal tract configuration only as much as is absolutely necessary a) that the utterance is understood and b) that the listener(s) do not notice. This is an extremely difficult task and requires high attention. Therefore, a quick way to find out whether articulatory avoidance has been performed is to compare the beginning (high attention) and the end (low attention) of the recording. In the case where the speaker articulates with higher precision and more phonological and phonetic consistency at end of the session<sup>124</sup>, i.e. when he or she is already tired and inattentive, articulatory avoidance can be assumed. Since attention is higher at the beginning of a recording session, under normal conditions less backgrounding processes are applied at the beginning (Vanecek & Dressler 1977, Moosmüller 1997d) as compared to the end of a recording session.

It has already been stated that in articulatory avoidance, the intended output target is the neutral vocal tract configuration. This is also performed by some speakers. The result is no change in the articulatory configuration as the speaker moves from phoneme to phoneme, and, consequently, there is also no change in formant frequencies over time. Changes in time can only be observed for the fundamental frequency. What is left is a sort of singsong which is thought to denote the sentence.

Not to articulate at all is of course a very salient method when it is one's aim to avoid articulation. It is rather unwise to expose no articulation at all, because this is of course immediately revealed by the listener. Therefore, a wiser strategy is to show some articulation, as in the performance of a speaker of Portuguese based Crioulo, as spoken in Guinea Bissau. The speaker is asked to repeat a list of sentences, and the list is read to him four times. So, a certain timespan has passed till he hears sentence 1 again. Figures 6.4 shows the spectrogram of the four utterances of the word "ntene" (I have) from the sequence "ntene centu i cincu" (I have hundred and five) from the four

---

<sup>124</sup> This holds only if the speaker does not know the end of the session. If he or she does, attention increases again at the end (Vanecek & Dressler 1977).

successive rounds. The successive items are separated by a timespan of approximately 15 minutes.

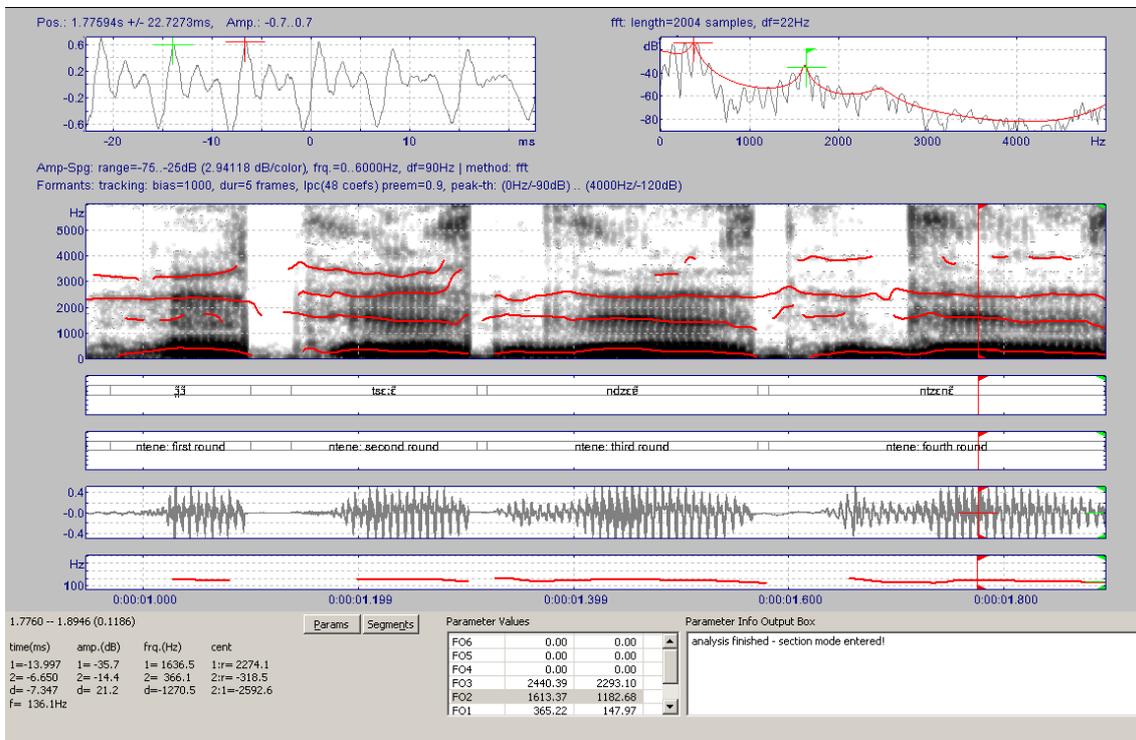


Figure 6.4: Spectrogram of four successive items of the word „ntene“ (I have), language: Crioulo, Guinea Bissau, male speaker. Bottom panel: Fundamental frequency, Next panel from bottom: waveform window, 3<sup>rd</sup> panel from bottom: segment names, 4<sup>th</sup> panel from bottom: phonetic transcription, left upper panel: waveform zoom window, right upper panel: amplitude spectrum window.

Figure 6.4. vividly shows that in each round, the item is articulated differently and that precision of articulation increases from item to item. This is also manifested in the duration of the items:

- ntene, 1<sup>st</sup> round: 136 ms
- ntene, 2<sup>nd</sup> round: 173 ms
- ntene, 3<sup>rd</sup> round: 252 ms
- ntene, 4<sup>th</sup> round: 313 ms

The first item<sup>125</sup> (136 ms) consists of a nasalized vowel, the quality of which is not easily definable. The whole sequence is nasalized. The first part is additionally articulated with breathy voice. There are also some discontinuities in the contour of the second formant, which might point to some coupling with the pharynx (Stevens 2003) and adds to the undefinability of the output.

The second item (with 173 ms) adds the plosive, which is affricated. The formant structure is clear and exposes some movement over time.

In the third item (with 252 ms) the personal pronoun /n/ “n” (I) is articulated. The sequence nasal consonant + plosive is fully voiced, i.e. the closure phase of the plosive is voiced as well, and the plosive is again affricated. The formant structure reveals that two vowels are articulated in sequence. The intervocalic nasal consonant is still missing.

In the spectrogram of the fourth item (with 313 ms), an indication of the intervocalic nasal consonant can be found. The initial nasal consonant, denoting the personal pronoun, is articulated as well, followed by a fortis plosive, which is voiced and affricated. Therefore, it is the last item of the fourth round, where it can be assumed that the speaker is already tired after approximately 90 minutes of recording, which exposes the most precise articulation, with each phoneme having an output representation.

It are these last two items which might occur in spontaneous speech as well. In spontaneous speech, the last syllable of “ntene” might be absorbed by the preceding vowel, supplying it with nasalization and resulting in [nt̃] (see Figure 6.5).

---

<sup>125</sup> All items expose a more or less strong frication at the end, which points to the start of the subsequent fricative /s/ of “centu” (hundred).

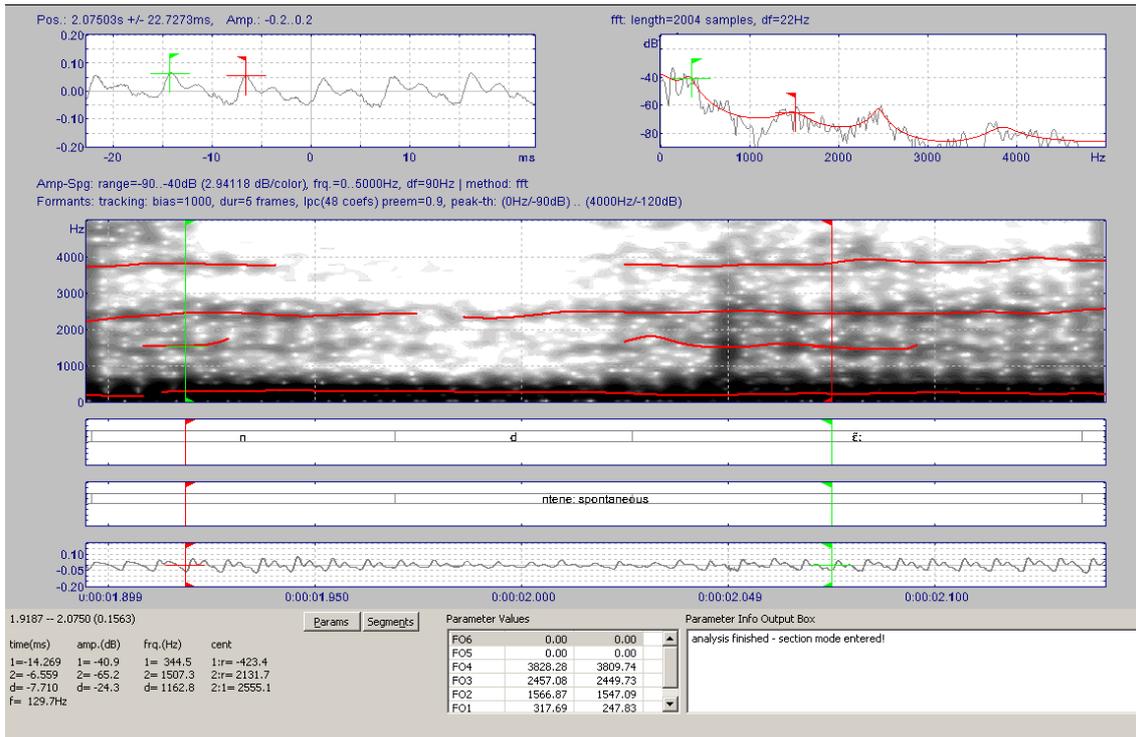


Figure 6.5: Spectrogram of the word „ntene“ (I have), spontaneous speech, language: Crioulo, Guinea Bissau, male speaker. Bottom panel: waveform window, next panel from bottom: segment names, 3<sup>rd</sup> panel from bottom: phonetic transcription, 4<sup>th</sup> panel from bottom: spectrogram, left upper panel: waveform zoom window, right upper panel: amplitude spectrum window.

Figure 6.5 exposes the most reduced form of „ntene“ found in the spontaneous speech of this speaker. Even in the most reduced form, the personal pronoun /n/ “n” (I) is articulated and the plosive is not deleted, not even fricated. Since Crioulo verb conjugation lacks personal agreement, the personal pronoun carries highly relevant information and is not supposed to be easily deleted, even in very reduced forms<sup>126</sup>. The other, less reduced forms of “ntene”, uttered in spontaneous speech, expose a clearer formant structure, a clearer closure phase, which is voiced, and a fortis plosive. One item even shows multiple releases.

<sup>126</sup> There are of course languages which have only non-obligatory constituents, such as Tibetan (Vollmann 2005).

Therefore, it can be observed that

- in spontaneous speech, the speaker exhibits a higher articulatory precision with respect to clear cut segment boundaries,
- in repeating utterances, the speaker exhibits the highest precision when it can be assumed that attention is already lower.
- Moreover, the first two items of “ntene”, produced in the repeating sentence task, display the application of processes which are not allowed in the spontaneous speech of the language (deletion of /n#/ and /n#/).

Process application takes place contrary to phonological, socio-, and psycho-phonological observations: in the most formal task, the repeating sentences task, the speaker applies more reduction processes than in the less formal task (spontaneous speech). He applies more reduction processes when attention is high as compared to a situation when attention is already low.

It is especially the lack of consistency in the production of the utterances which points to articulatory avoidance. Since the target is very imprecisely defined as “only depart from the neutral vocal tract configuration as much as is absolutely necessary to give the impression of a properly pronounced segment”, such an imprecisely defined target is constantly missed. This results in inconsistent and arbitrary application and suppression of phonological and phonetic processes, which are in the discussed example:

- deletion of segments which are not allowed to be deleted in the language and
- indefinable vowel quality in the first item.

The discussed cases of undershoot show that undershoot is not something that happens unnoticed in “normal” speech production, but that it is perceived both by the speaker and by the listener. There are many more cases where undershoot can happen to a speaker (extreme tiredness, unattentiveness, drunkenness or being under the influence of

drugs, motor speech disorders), however, the result is in any case unsatisfactory both for the speaker and the listener.

## 6.5. Another look at invariance

Invariance depends on the linguistic and extralinguistic context. This approach differs from Keating (1990) and Guenther<sup>127</sup> (1995, 2003) insofar as the phoneme is conceptualized as invariant, whereas the target (goal) can take different shapes. The phoneme is modeled via phonological and phonetic processes, which are applied or suppressed according to surrounding segments, and prosodic, sociolinguistic, and psycholinguistic circumstances.

The target is variant insofar, as it might change its appearance depending on context<sup>128</sup>. Variance of the target is guided by processes which, in any case, are phonetically motivated and planned. It is the process which makes the target invariant. Instead of conceptualizing wide and narrow windows which introduce some arbitrariness exactly because they contain all phonetic information, the way from phoneme to target is mediated by processes which are applied or not<sup>129</sup>, in dependence on linguistic and extralinguistic factors.

This means, the clearer the context is defined, the less variability will appear. In the following section, variability of the three vowels<sup>130</sup> /i, e, a/ under two conditions was compared. In the first condition (Condition 1), speakers were asked to repeat one and the same sentence until they were told to stop (after the tenth time). In this

---

<sup>127</sup> Guenther, similar to Keating, defines targets as ranges: “These targets are defined in a planning space made up of auditory and orosensory dimensions. For example, the target for vowel sounds specifies a range of acceptable values of formant ratios.” Guenther 2003: 214)

<sup>128</sup> Context is defined here in a wide sense, comprising both linguistic and extralinguistic factors.

<sup>129</sup> The variability of application vs. suppression of processes is not included in Keating’s model.

<sup>130</sup> This repeating sentences task was originally designed within another project for another purpose, therefore, not all vowels could be tested.

condition, the context was maximally restricted. Under the second condition (Condition 2: a reading task), vowels were only controlled for phonetic context and two stress levels (stressed vs. unstressed), but not for sentence position. For each speaker, the variability coefficient was calculated for F1, F2, and F3 for each vowel of each condition at vowel midpoint, where the vowel is supposed to expose a stationary part, and for the whole vowel. One-tailed t-tests were performed in order to test whether variability differed according to conditions.

In Table 6.6 the results for the vowels in stressed position, calculated at vowel midpoint, are presented.

Var. coeff.	C1	C2	C1	C2	C1	C2
<b>stressed</b>	/i/	/i/	/e/	/e/	/a/	/a/
F1	11.77	9.84	4.90	6.74	14.74	10.37
F2	2.14	2.43	<b>2.29</b>	<b>3.79</b>	5.03	5.50
F3	<b>4.53</b>	<b>7.80</b>	<b>3.09</b>	<b>6.15</b>	<b>2.89</b>	<b>4.97</b>

Table 6.6: Variability coefficients calculated at vowel midpoint under two different conditions (C). C1: vowel drawn from one and the same sentence repeated ten times, c2: vowel controlled for phonetic context. Statistically significant differences ( $p \leq 0.05$ ) are printed in bold.

It can be read from Table 6.6 that, whenever statistically significant differences appear, the variability coefficient is higher in the Condition 2 as compared to the Condition 1. It is most conspicuous that for all three vowels, F3 exposes statistically significant differences, i.e. on such a fine graded level, F3 becomes more important and balanced. The vowel /e/ exposes differences for F2 as well. To give an example, three successive frames of all /e/ vowels calculated at vowel midpoint for six speakers are presented in Figure 6.6 (for Condition 2) and Figure 6.7 (for Condition 1).

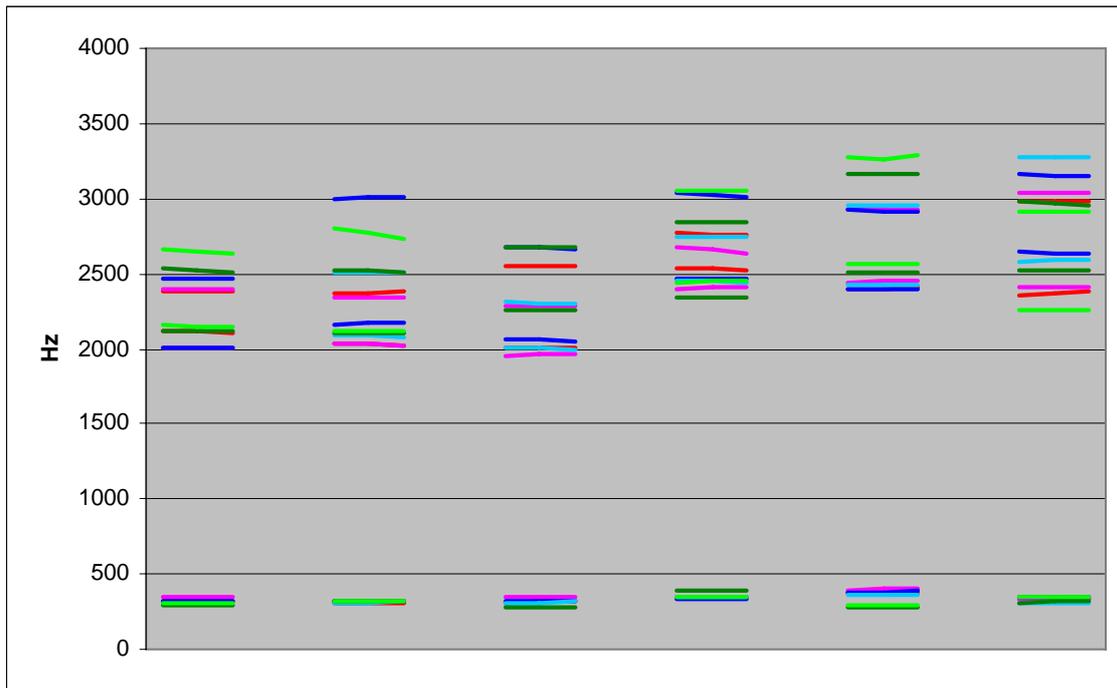


Figure 6.6: Three successive frames of F1, F2, F3 measured at vowel midpoint of the stressed vowel /e/. Reddish lines: “geben” (to give), bluish lines: “gegeben” (have given), greenish lines: “vergeblich” (in vain), two repetitions of each item. Column 1 to 3: 3 male speakers, column 4 to 6: 3 female speakers.

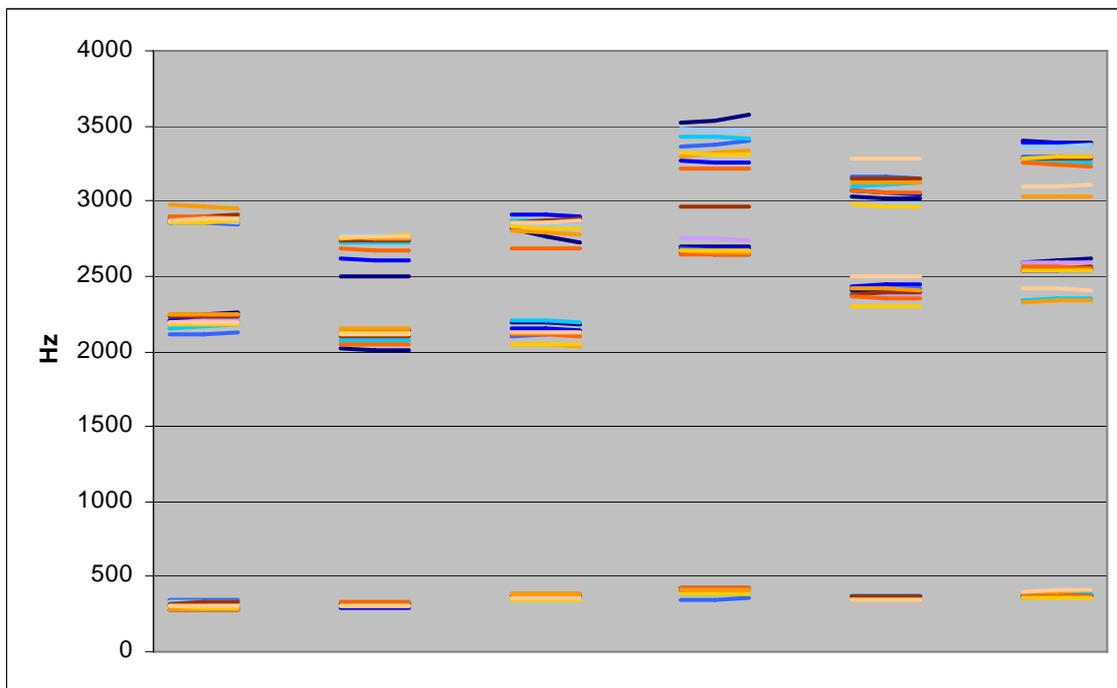


Figure 6.7: Three successive frames of F1, F2, and F3 measured at vowel midpoint of the stressed vowel /e/ from “Leber” (liver). Bluish lines: repetition 1-5, brownish lines: repetition 6-10. Column 1-3: 3 male speakers, column 4-6: 3 female speakers.

A visual comparison of Figure 6.6 and Figure 6.7 shows not only that the vertical span is lower within each formant in Figure 6.7, but also, that a clear demarcation between F2 and F3 is apparent, whereas in Figure 6.6, this demarcation between F2 and F3 is partly smeared. The blurring in Figure 6.6 results from the fact that although vowels all carry primary lexical stress and share the same phonetic context, vowels from different prosodic positions have been superimposed. This means that ultimately, invariance lies not so much in the phonetic context than in precisely determined stress assignment which attributes the ultimate shape to the phoneme (see 6.6.3).

In the unstressed position, statistically significant differences between the conditions appear only for F3 of the vowel /i/ (see Table 6.7):

Var. coeff.	C1	C2	C1	C2	C1	C2
unstressed	/i/	/i/	/e/	/e/	/a/	/a/
F1	7.70	8.45	6.74	6.00	17.93	8.91
F2	3.28	4.46	4.95	4.43	4.44	5.67
F3	<b>3.42</b>	<b>5.73</b>	3.62	3.43	2.52	3.29

Table 6.7: Variability coefficients calculated at vowel midpoint under two different conditions (C). C1: vowel drawn from one and the same sentence repeated ten times, C2: vowel controlled for phonetic context. Statistically significant differences ( $p \leq 0.05$ ) are printed in bold.

It has already been explicated that the vowel /i/ is articulated in an acoustically unstable region which results in a high variability (especially of F3). This becomes apparent here again, where F3 shows more variability in Condition 2 as compared to Condition 1. The other vowels exhibit no statistically significant differences between the conditions. These results show, moreover, that variability as a whole is reduced in unstressed positions. It has already been put forward in Moosmüller (2002), and will be dealt with in 6.6.2, that unstressed positions exhibit less variability than stressed positions. This is confirmed here insofar as no statistically significant differences appear for the vowels /e/ and /a/.

More variability in the stressed position, however, does not mean that articulation is less precise in stressed positions. On the contrary, it has to be interpreted in the way

that more information is given by finely tuning several levels of stressed positions. The weakest prosodic positions are not differentiated that much (see 6.6.3).

“Now in languages with “phonemic stress” ( $\approx$  stress-timed languages) synchronic obscuration processes are maximized in unstressed syllables (and syllable-finally), clarification processes are maximized in stressed syllables (and word-initially), thus allowing more sounds in stressed than in unstressed syllables.” (Dressler 1979: 268).

## 6.6. Prosodic analysis of variability

In 6.5, it has been argued that identical context and identical sentence/utterance position of a given segment bears some invariance, insofar as the quality of a given segment experiences less variability than phonologically identical segments in different sentence/utterance positions. So far, little has been said about the quality of the changes and why they occur.

### 6.6.1. Duration

There is general agreement that prosodic strength is responsible for the qualitative changes of phonologically identical segments. Phonemes in stressed positions are articulated differently from the same phonemes in unstressed positions. Since unstressed positions experience durational shortenings as well<sup>131</sup>, the changes observed in unstressed positions are usually termed “reduction”. As compared to stressed positions, unstressed positions are shorter in Standard Austrian German. Figure 6.8 shows the distribution of the durations, measured in number of periods (NoP), for all stressed and unstressed vowels in the sentence reading task of speaker sp082:

---

<sup>131</sup> Crystal & House (1988a) found a difference of more than 60 ms for stressed vs. unstressed vowels in American English. However, Podevska & Adisasmito-Smith (1999) observed longer durations in unstressed positions for two speakers of Buginese, which they attributed to final lengthening.

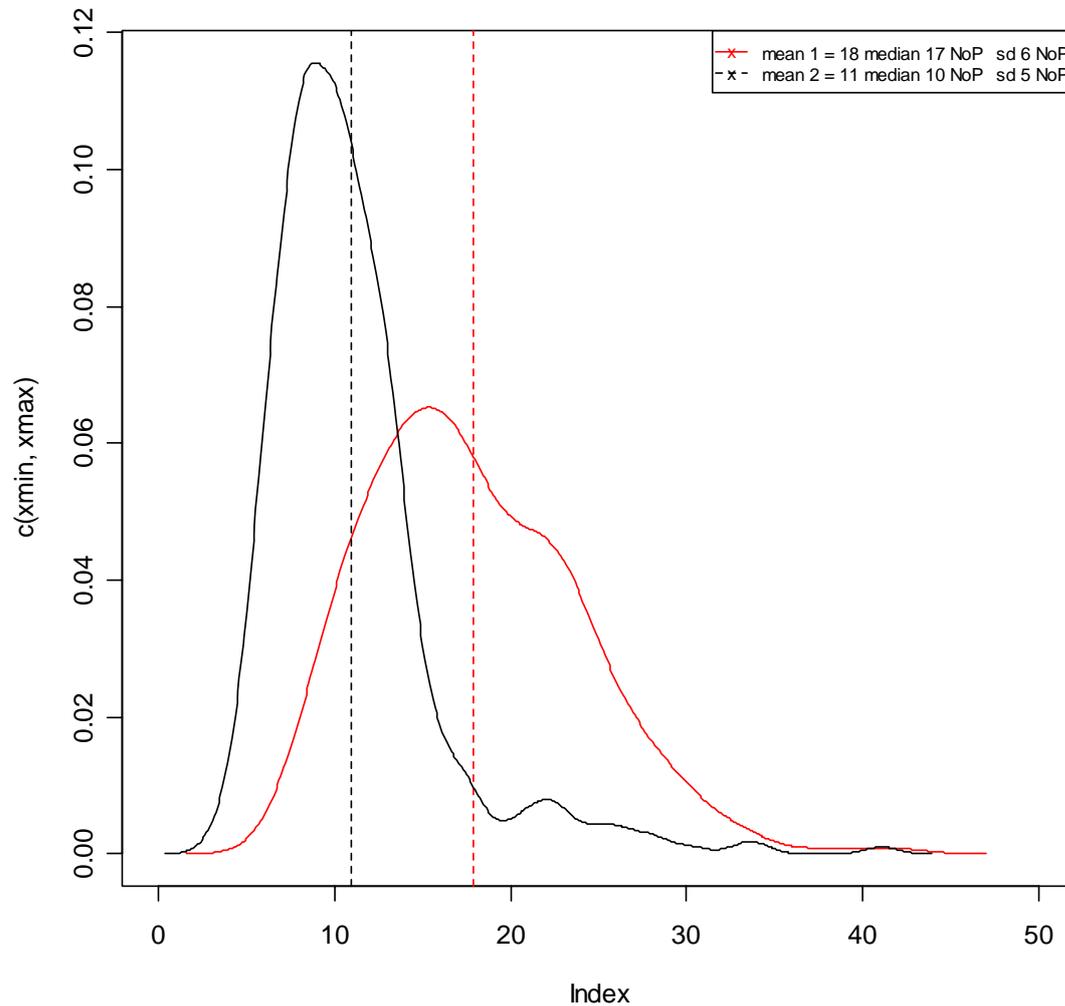


Figure 6.8: Density plot of the duration measurements (in NoP) of all vowels ( $n = 737$ ) in stressed (red) and unstressed (black) position of the sentence reading task, speaker sp082.

The results are statistically significant and hold for all speakers in both speaking tasks. Table 6.8 gives an overview of the mean NoP values and the statistical results of the one-tailed t-tests:

NoP	stressed	unstressed	t	p	stressed	unstressed	t	p
Sp180	18	12	13.71	0.00	14	10	8.19	0.00
Sp129	15	8	16.66	0.00	14	10	10.01	0.00
Sp082	18	11	16.96	0.00	17	11	7.68	0.00
Sp012	11	6	16.92	0.00	7	5	4.85	0.00
Sp126	13	7	15.69	0.00	12	7	9.32	0.00
Sp127	14	7	20.86	0.00	11	6	11.89	0.00

Table 6.8: Mean number of periods measured in stressed and unstressed positions, t-values and level of significance. The four columns to the left refer to the sentence reading task, the 4 columns to the right to spontaneous speech.

Table 6.8 shows that, for each speaker and within each speaking task, the vowels in unstressed positions are significantly shorter than the vowels in stressed positions. As concerns the two speaking tasks (sentence reading task and spontaneous speech), a comparison of the mean values also indicates a difference in the direction, and that the number of periods is smaller in spontaneous speech. However, statistically significant results could only be obtained for speakers sp012, sp127, sp129, and sp180. For speaker sp126 and speaker sp082, the differences did not prove to be statistically significant, both for the comparison of stressed and unstressed position.

These results strongly suggest that changes in vowel quality are a result of durational truncation. Tables 6.9 to 6.11 give the correlation coefficient of NoP and F1, F2, or F3.

r: NoP/F1	Sp180	Sp129	Sp082	Sp012	Sp126	Sp127
/i/	-0.18	-0.40	-0.16	-0.06	-0.15	<b>-0.47</b>
/ɪ/	-0.09	0.09	-0.27	-0.36	0.08	-0.01
/y/	-0.40	<b>-0.97</b>	<b>-0.86</b>	-0.26	<b>-0.65</b>	<b>-0.95</b>
/ʏ/	0.09	-0.21	0.39	-0.52	-0.30	<b>-0.69</b>
/e/	<b>0.40</b>	<b>-0.44</b>	<b>0.43</b>	-0.27	<b>-0.39</b>	-0.06
/ɛ/	0.22	-0.12	0.26	0.07	0.19	-0.21
/ø/	0.34	-0.22	0.50	-0.26	0.30	0.44
/œ/	<b>0.71</b>	-0.52	-0.15	0.57	-0.02	-0.39
/u/	-0.18	-0.22	<b>-0.60</b>	0.18	-0.11	-0.64
/ʊ/	-0.36	-0.22	0.36	-0.02	0.21	0.24
/o/	0.16	-0.18	0.34	0.36	0.20	0.16
/ɔ/	-0.01	0.18	-0.08	-0.06	0.05	-0.10

/a/	0.20	0.15	0.25	0.21	<b>0.29</b>	<b>0.29</b>
/i/	<b>-0.44</b>	0.14	-0.01	-0.06	-0.13	-0.21
/ɪ/	-0.09	<b>0.41</b>	0.00	-0.06	<b>0.31</b>	0.12
/y/	-0.21	-0.09	<b>0.72</b>	-0.08	-0.62	-0.08
/ʏ/	–	0.23	–	–	–	<b>0.67</b>
/e/	0.21	0.33	<b>0.55</b>	-0.09	0.28	0.47
/ɛ/	0.17	<b>0.18</b>	0.48	0.13	<b>0.52</b>	<b>0.25</b>
/u/	-0.06	-0.02	<b>0.60</b>	-0.22	-0.13	0.11
/ʊ/	0.18	0.02	0.86	0.30	0.07	-0.07
/o/	-0.24	0.35	0.06	-0.42	<b>0.78</b>	-0.07
/ɔ/	0.38	0.23	0.14	-0.27	0.02	0.29
/ɑ/	0.25	0.16	<b>0.42</b>	<b>0.41</b>	<b>0.30</b>	<b>0.29</b>

Table 6.9:  $r$  for duration (number of periods)/F1 in stressed position (first six rows) and in unstressed position (last six rows), broken for all speakers, sentence reading task, statistically significant results ( $p \leq 0.05$ ) are in bold. Where statistically significant  $r$ -values have the opposite sign to the hypothetical result, the value is additionally in italics.

r: NoP/F2	Sp180	Sp129	Sp082	Sp012	Sp126	Sp127
/i/	-0.10	0.11	<b>0.50</b>	<b>0.45</b>	<b>0.65</b>	<b>0.57</b>
/ɪ/	0.04	-0.01	0.24	0.16	0.19	0.04
/y/	-0.02	-0.45	0.49	0.17	0.26	0.22
/ʏ/	-0.17	-0.02	-0.21	<b>0.76</b>	0.69	<b>0.70</b>
/e/	0.34	<b>0.53</b>	0.23	0.31	<b>0.57</b>	<b>0.38</b>
/ɛ/	0.06	0.04	-0.02	-0.03	-0.02	<b>0.35</b>
/ø/	<b>-0.73</b>	-0.14	<b>-0.61</b>	-0.43	-0.45	0.43
/œ/	-0.32	<b>0.76</b>	0.13	-0.22	0.35	0.27
/u/	-0.43	-0.31	-0.67	-0.25	-0.26	-0.04
/ʊ/	-0.07	-0.28	0.33	0.28	0.49	-0.37
/o/	-0.20	-0.03	0.19	-0.11	-0.04	0.16
/ɔ/	-0.21	-0.05	<b>-0.57</b>	-0.23	-0.01	-0.09
/ɑ/	0.19	<b>0.45</b>	-0.05	-0.02	0.09	<b>0.27</b>
/i/	<b>0.51</b>	-0.04	0.05	0.19	0.01	-0.07
/ɪ/	-0.07	0.08	-0.13	-0.10	-0.12	0.14
/y/	0.36	-0.21	0.43	0.57	0.70	0.48
/ʏ/	–	<b>0.82</b>	–	–	–	-0.10
/e/	-0.17	-0.10	-0.16	0.25	0.03	-0.06
/ɛ/	-0.07	0.10	0.02	0.07	0.10	0.09
/u/	-0.18	-0.27	0.37	-0.03	-0.17	-0.02
/ʊ/	0.23	-0.32	0.38	<b>-0.93</b>	0.09	0.01

/o/	-0.01	0.35	-0.55	-0.20	<b>0.61</b>	0.16
/ɔ/	0.05	0.11	0.33	0.12	<b>0.57</b>	<b>0.54</b>
/ɑ/	-0.18	-0.04	-0.21	-0.28	-0.13	-0.23

Table 6.10:  $r$  for duration (number of periods)/F2 in stressed position (first six rows) and in unstressed position (last six rows), broken for all speakers, sentence reading task, statistically significant results ( $p \leq 0.05$ ) are in bold. Where statistically significant  $r$ -values have the opposite sign to the hypothetical result, the value is additionally in italics.

r: NoP/F3	Sp180	Sp129	Sp082	Sp012	Sp126	Sp127
/i/	0.12	<b>0.34</b>	<b>0.50</b>	<b>0.56</b>	<b>0.66</b>	<b>0.75</b>
/ɪ/	0.13	0.06	0.18	0.17	0.25	0.04
/y/	0.58	<b>-0.95</b>	0.15	-0.48	-0.33	-0.51
/ʏ/	-0.10	0.16	0.45	0.52	-0.06	-0.38
/e/	0.00	<b>0.62</b>	0.29	<b>0.49</b>	0.41	<b>0.39</b>
/ɛ/	<b>0.33</b>	0.15	0.25	-0.13	0.17	<b>0.35</b>
/ø/	-0.34	-0.04	0.08	<b>-0.87</b>	0.06	-0.06
/œ/	0.11	0.44	-0.46	-0.19	0.14	0.40
/u/	0.20	-0.04	-0.07	0.06	0.33	-0.14
/ʊ/	0.47	<b>-0.64</b>	0.47	-0.34	-0.39	0.26
/o/	0.02	0.18	0.37	0.02	-0.36	-0.00
/ɔ/	0.16	-0.32	0.31	0.06	0.39	<b>0.47</b>
/ɑ/	0.22	0.45	<b>0.48</b>	<b>0.28</b>	<b>0.29</b>	<b>0.32</b>
/i/	<b>0.47</b>	0.03	0.05	0.21	0.08	-0.21
/ɪ/	0.06	0.04	0.08	-0.02	-0.11	<b>0.31</b>
/y/	0.01	-0.02	-0.20	<b>0.80</b>	-0.33	0.27
/ʏ/	–	<b>0.72</b>	–	–	–	-0.47
/e/	0.11	0.07	0.03	-0.16	0.04	0.15
/ɛ/	0.12	0.02	0.09	-0.08	0.34	0.13
/u/	-0.01	<b>-0.52</b>	0.05	0.39	0.27	-0.30
/ʊ/	-0.01	-0.34	<b>0.94</b>	0.44	-0.69	0.30
/o/	-0.03	0.13	0.30	0.49	-0.11	<b>0.64</b>
/ɔ/	-0.03	-0.12	-0.17	-0.25	-0.10	0.52
/ɑ/	0.08	-0.09	-0.17	<b>-0.36</b>	0.15	-0.08

Table 6.11:  $r$  for duration (number of periods)/F3 in stressed position (first six rows) and in unstressed position (last six rows), broken for all speakers, sentence reading task, statistically significant results ( $p \leq 0.05$ ) are in bold. Where statistically significant  $r$ -values have the opposite sign to the hypothetical result, the value is additionally in italics.

Tables 6.9 to 6.11 demonstrate that statistically significant correlations turn up only sporadically. Moreover, in unstressed positions, the observed correlations often indicate a direction contrary to theoretical assumptions (e.g. the shorter the duration for /u/, the lower F1). Quite often, the observed correlation is not very strong either. Some trends for a correlation of duration and F2 and F3 can be observed with the male speakers (sp012, sp126, sp127) for the vowel /i/, the strongest holding for speaker sp127. From these results it can be concluded that duration plays no relevant role in the qualitative change of vowels. These results are in accordance with other studies (Gay 1978, Fourakis 1991, van Son & Pols 1990, 1992, but see Nowak 2006 for contradictory results on Polish) that could not prove a correlation of duration and change in vowel quality.

### 6.6.2. Stress<sup>132</sup>

As has been observed by many studies so far (see e.g. Gay 1978, Nord 1986, Dogil & Williams 1999, Erickson 2002, Wouters & Macon 2002, van Son & Pols 2002, Padgett & Tabain 2005, to name just a few), prosodic strength is a highly relevant factor in determining the quality of a vowel in Standard Austrian German. Already, by dividing the analysed vowels into only two prosodic strengths (stressed and unstressed)<sup>133</sup>, statistically significant differences occur between vowels in stressed and unstressed positions, both in the sentence reading task and in spontaneous speech. For the majority of the data, stressed and unstressed vowels are at least discriminated by one of the three lowest formants<sup>134</sup>.

---

<sup>132</sup> It has to be emphasized that in the chapters on stress (6.6.2, 6.6.3, 6.6.4) and on rhythm (6.6.5), only the contribution of the spectral information of the vowels (F1, F2, F3) to stress and rhythm has been investigated, since the contribution of prosodic parameters (F0, duration, intensity) is not the main objective of this research.

<sup>133</sup> For this coarse division, stress was assigned by reference to lexical stress for content words and function words were labeled as unstressed throughout.

<sup>134</sup> Speaker sp126 exposed no differences between stressed and unstressed /u/ and /o/, and speaker sp180 showed no differences between stressed and unstressed /ʊ/.

In Figure 6.9, the overall results of the differences between stressed and unstressed vowels are presented. For each formant which showed a statistically significant difference between unstressed and stressed vowels of a given category, the difference (in %) was calculated per speaker. Results were root-squared, in order to eliminate negative values. Per vowel and speaker, the statistically significant differences of F1, F2, and F3 were added up. One-tailed t-tests were calculated in order to find out whether the discriminatory power is greater in the sentence reading task as compared to spontaneous speech. For most vowels, the discriminatory power of the sentence reading task did not exceed the discriminatory power of spontaneous speech. The majority of the vowels showed no statistically significant differences. The discriminatory power of the sentence reading task is greater for the vowels /e/ and /u/, whilst the discriminatory power of spontaneous speech is greater for the vowel /i/ (see Figure 6.9).

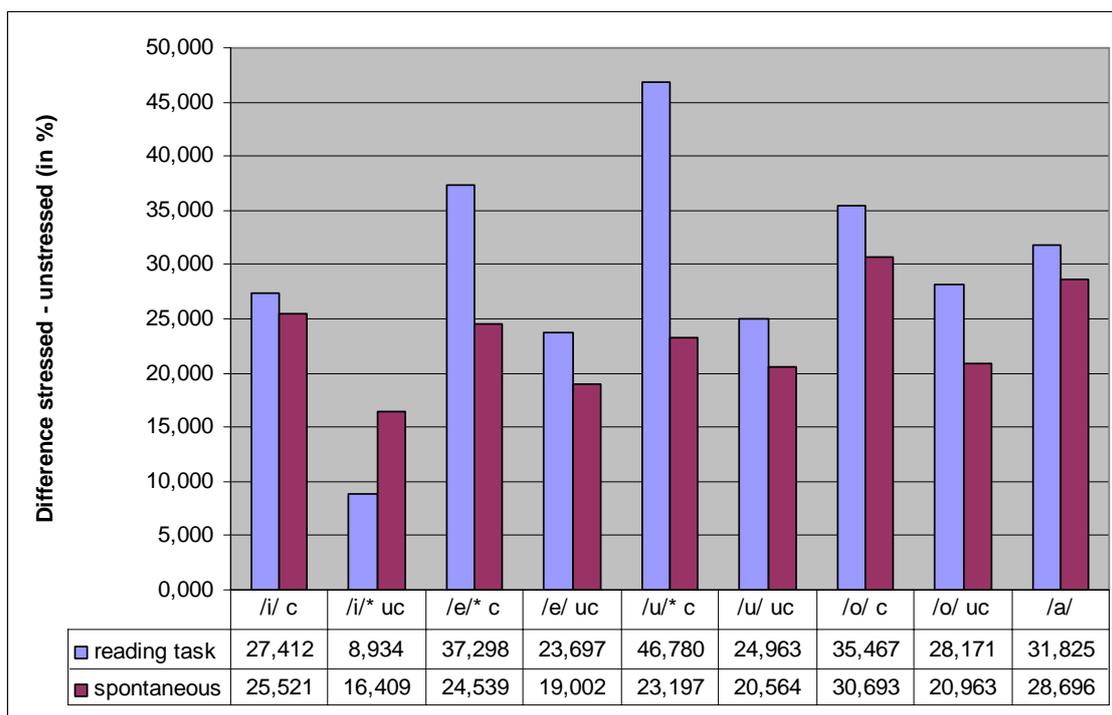


Figure 6.9: Sum of the statistically significant differences (in %) of F1, F2, and F3 between stressed and unstressed vowels, pooled over all speakers. The asterisk indicates a statistically significant difference ( $p < 0.02$ ) between the two speaking tasks. Legend: c = [+constricted], uc = [-constricted].

Although the overall results only show little differences between the two speaking tasks, the way this overall pattern is achieved can differ considerably. For example, in the sentence reading task, statistically significant differences between stressed and unstressed /i/ show up for F1, F2, and F3, whereas, in spontaneous speech, only two speakers (sp126 and sp180) differentiate stressed and unstressed /i/ for F1. The missing F1 differentiation is compensated by a higher difference in both F2 and F3 in spontaneous speech, so that, in the end, both speaking tasks arrive at the same overall discriminatory power. The compensatory effect is missing in the vowels /e/ and /u/, which have the highest discriminatory power in the sentence reading task, i.e, they show the greatest changes in vowel quality.

Stressed and unstressed vowels are primarily discriminated by F2. The least discriminatory power is exerted by F3. This result is statistically significant ( $p < 0.01$ ). Broken down for vowel category, the following picture emerges (see Figure 6.10):

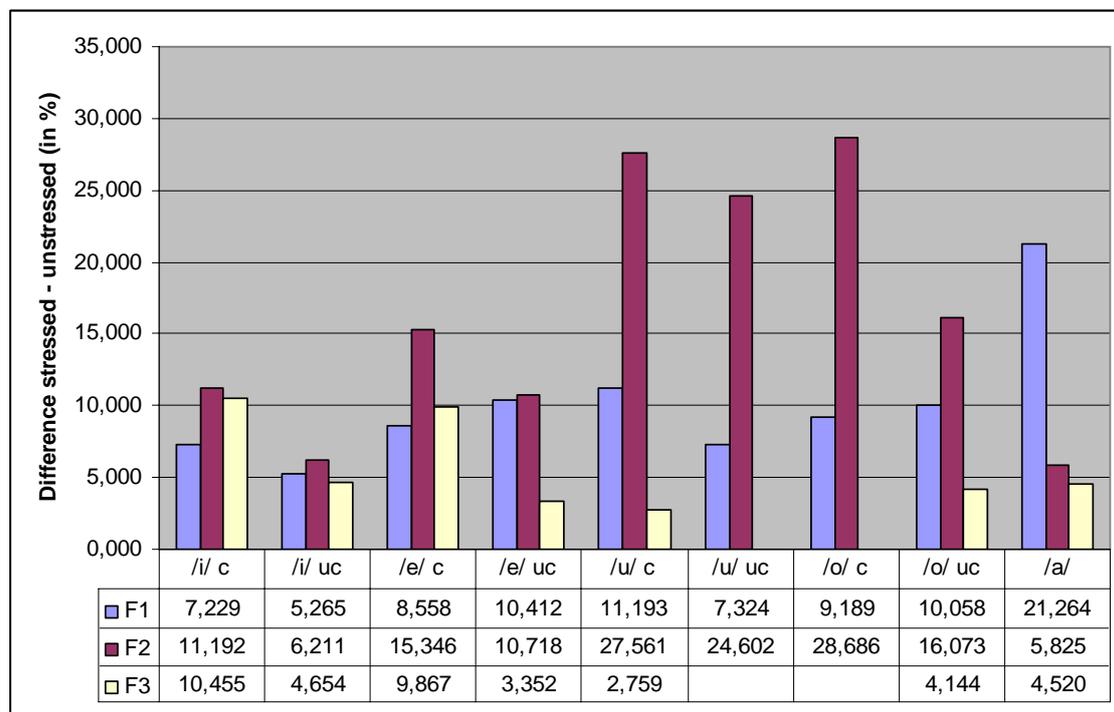


Figure 6.10: Mean values of the statistically significant differences (in %) between stressed and unstressed vowels, broken for F1, F2, and F3, pooled over all speakers and speaking tasks. Legend: c = [+constricted], uc = [-constricted].

It becomes apparent from Figure 6.10 that, for almost every vowel category, F2 plays the dominant role in differentiating stressed from unstressed vowels. A salient exception is the vowel /a/, which discriminates stressed from unstressed vowels predominantly via F1 (i.e. the degree of lip opening is smaller for the unstressed /a/s). It is worth mentioning that for the [–constricted] vowels /ɛ/ and /ɔ/ – for those speakers where statistically significant differences appear<sup>135</sup> – F1 is lowered in the unstressed position as well (i.e. the degree of lip opening decreases in unstressed position). For the other vowels where statistically significant differences appear, F1 becomes higher in the unstressed position. These results suggest that F1 varies less in unstressed positions as compared to stressed positions. The calculation of the variability coefficient over all stressed and unstressed vowels corroborates this assumption: the variability of F1 of the stressed vowels is higher than the variability of the unstressed vowels ( $p < 0.01$ )<sup>136</sup>. Figure 6.11 shows the results of the calculated variability coefficient over all vowels for F1 in spontaneous speech:

---

<sup>135</sup> In the sentence reading task, all speakers exhibited a statistically significant difference of F1 of the vowels /ɛ/ and /ɔ/. In spontaneous speech, statistically significant differences could only be observed for the speakers sp126, sp127, sp082, and sp180 for the vowel /ɛ/, and the speakers sp082 and sp180 for the vowel /ɔ/.

<sup>136</sup> One-tailed t-tests have been performed.

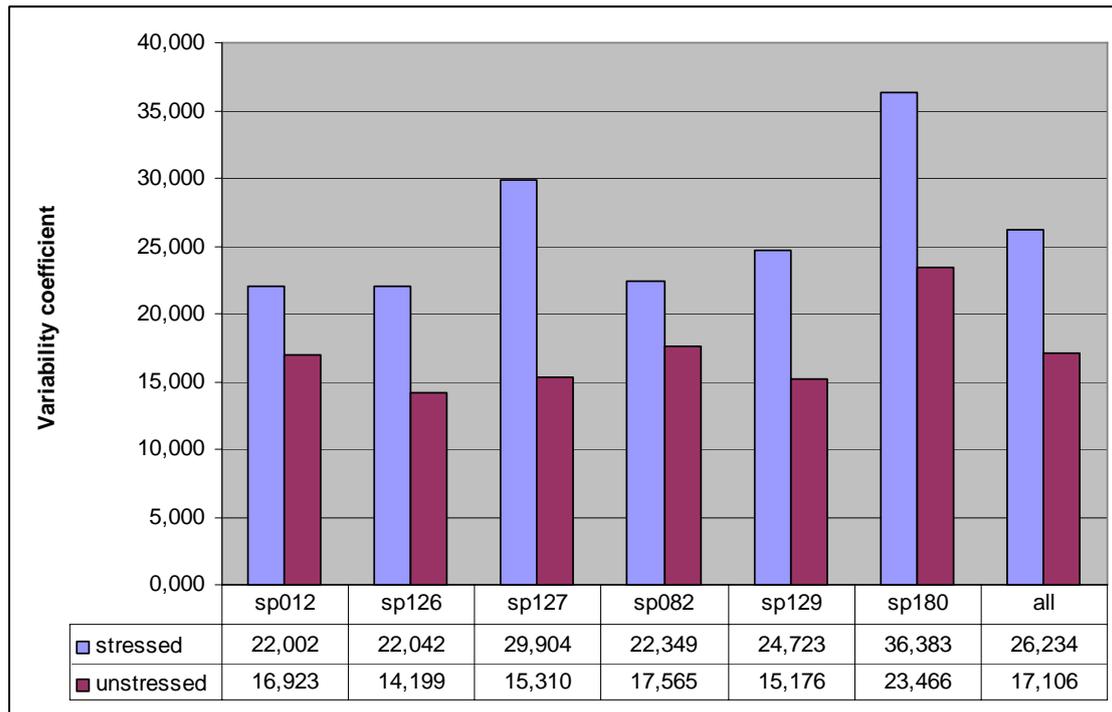


Figure 6.11: Variability coefficient of F1 calculated over all vowels in stressed and unstressed position, broken for speakers, spontaneous speech. Legend: sp = speaker.

The results for the sentence reading task do not differ from spontaneous speech. The calculated variability coefficient is about the same: 28.24 in stressed position and 16.47 in unstressed position (pooled over all speakers). Consequently, no statistically significant differences appear between the two speaking tasks, either in stressed, or in unstressed position.

F2, in a case where statistically significant difference occurs between stressed and unstressed vowels, is lowered for the front vowels and raised for the back vowels, due to a reduction of constriction degree, constriction length and lip protrusion (for the back rounded vowels). The changes involved with respect to F2 lead to a substantial and statistically significant ( $p < 0.01$ ) reduction of the variability of unstressed vowels. Figure 6.12 gives the variability coefficient calculated over all vowels in stressed and unstressed positions in spontaneous speech.

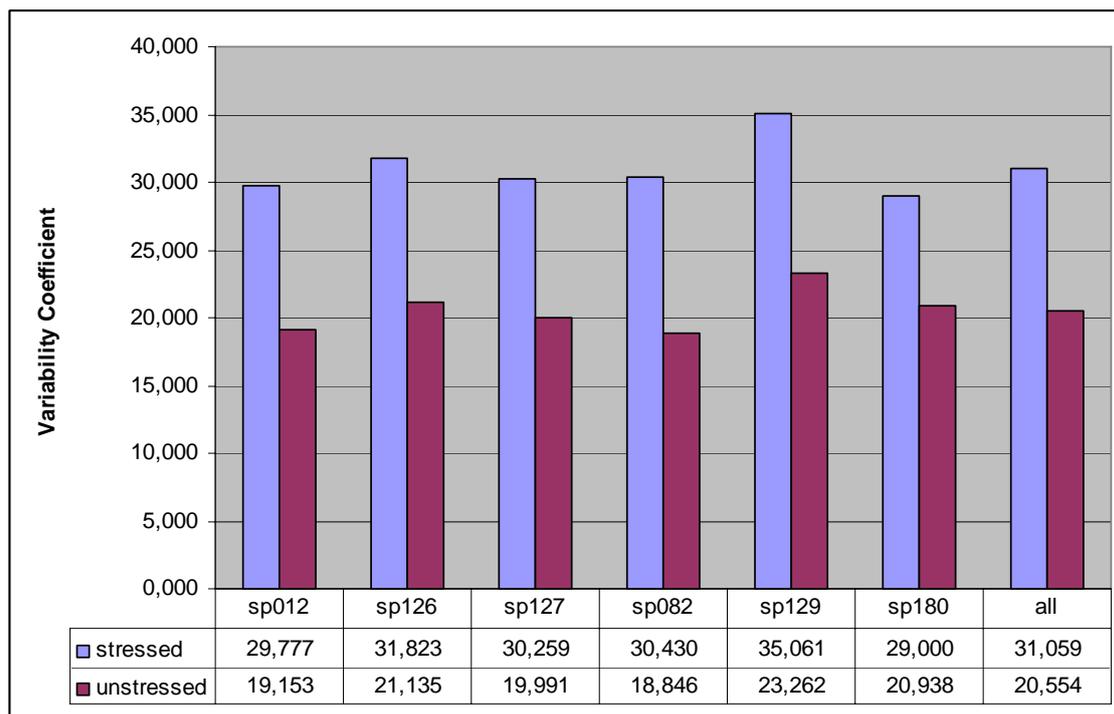


Figure 6.12: Variability coefficient of F2 calculated for all vowels in stressed and unstressed positions, broken for speakers, spontaneous speech. Legend: sp = speaker.

The pattern is the same in the sentence reading task. However, in stressed positions, the calculated variability coefficient is higher than in spontaneous speech (35.66 vs. 31.06 respectively,  $p = 0.01$ ), whereas in unstressed positions, no differences occur between the two speaking tasks (20.49 in the sentence reading task and 20.55 in spontaneous speech,  $p = 0.49$ ). These findings are not surprising, and corroborate the assumption already put forward in Dressler (1979: 268) that more variability is to be expected in stressed positions.

### 6.6.3. Secondary stress

From the analysis presented in 6.6.2, it becomes evident that in Standard Austrian German, unstressed vowels are differentiated from stressed vowels by a change in vowel quality<sup>137</sup>. These results suggest that the particular stress assigned to a given

<sup>137</sup> But not in vowel category! Similar results for Standard German have been obtained by Dogil & Williams (1999) and Kleber & Klippahn (2006).

syllable/vowel determines the ultimate shape of that vowel to a high degree. In German word formation, up to four stress levels can be distinguished (Wurzel 1980)<sup>138</sup>. However, whereas in Standard German the main acoustic correlates for primary stress are duration (Goldbeck & Sendlmeier 1988, Jessen et al. 1995, Mengel 1997, Dogil & Williams 1999) and spectral tilt (especially skewness, Claßen et al. 1998), acoustic correlates of secondary stress could not be proved<sup>139</sup> (Mengel 2000, Kleber & Klippahn 2006<sup>140</sup>). It has therefore been suggested that secondary stress is solely a perceptual phenomenon (Mengel 2000), with the listener expecting a secondary stress at time intervals of approximately 300 ms (Schreuder 2006).

However, since vowels change their shape due to stress, secondary stress might also be indicated by a specific vowel quality which differs from both primary-stress vowels and unstressed vowels. It can be assumed that secondary-stress vowels stand between primary stressed vowels and unstressed vowels, differing in the degree of lip opening, the degree of constriction, the length of constriction, and the degree of lip protrusion. This hypothesis has been tested on words bearing secondary stress, for example “Aussaat” (sowing) or “Finanzminister” (finance minister), in both speaking tasks. For each speaker, one-tailed t-tests have been performed. Only those vowels were analysed for which enough items could be observed to make a statistical analysis meaningful.

---

<sup>138</sup> See Zonneveld et al. (1999) for word stress assignment in German (subchapter by Jessen) and Doleschal (1988) for a discussion of stress assignment in German compounds.

<sup>139</sup> Jessen (1993) deduces a secondary stress from the fact that the “tense – lax” opposition is preserved to a higher degree in pre-stress 2 (i.e. two positions before primary stress) positions as compared to the pre-stress 1 (one position before primary stress) positions, but to a lower degree in pre-stress 2 positions as compared to stressed positions. This approach rests on a different problem formulation (do “tense” and “lax” vowels neutralize under certain stress conditions?) and can therefore be compared neither with Kleber & Klippahn (2006) nor with the current investigation.

<sup>140</sup> For secondary stress, Kleber & Klippahn (2006) only analysed syllables in pre-stress 2 positions, as for example in: “Medziner” (physician, the analysed pre-stress 2 vowel is underlined), whereas in the current investigation, secondary stresses resulting from composition, for example in “Neben,fach” (minor field of study), or from stress preservation in morphologically complex words, for example in “aufge,wachsen” (grow up: PP), were analysed.

### 6.6.3.1 The vowel /a/

For the vowel /a/ it can be assumed that for vowels bearing secondary stress, the degree of lip opening is smaller as compared to the primary stressed vowels, but larger as compared to the unstressed vowels, resulting in a lower F1 as compared to the primary stressed vowels, but a higher F1 as compared to the unstressed vowels. Additionally, the degree of constriction might be less tight as compared to the primary stressed vowels, resulting in a higher F2 and F3 for the secondary stressed vowels. The following schematized change in formant frequency can be set up:

- F1 primary stress > F1 secondary stress > F1 unstressed
- F2 primary stress < F2 secondary stress < F2 unstressed
- F3 primary stress < F3 secondary stress < F3 unstressed

In Figure 6.13 the results of the one-tailed t-tests are summarized:

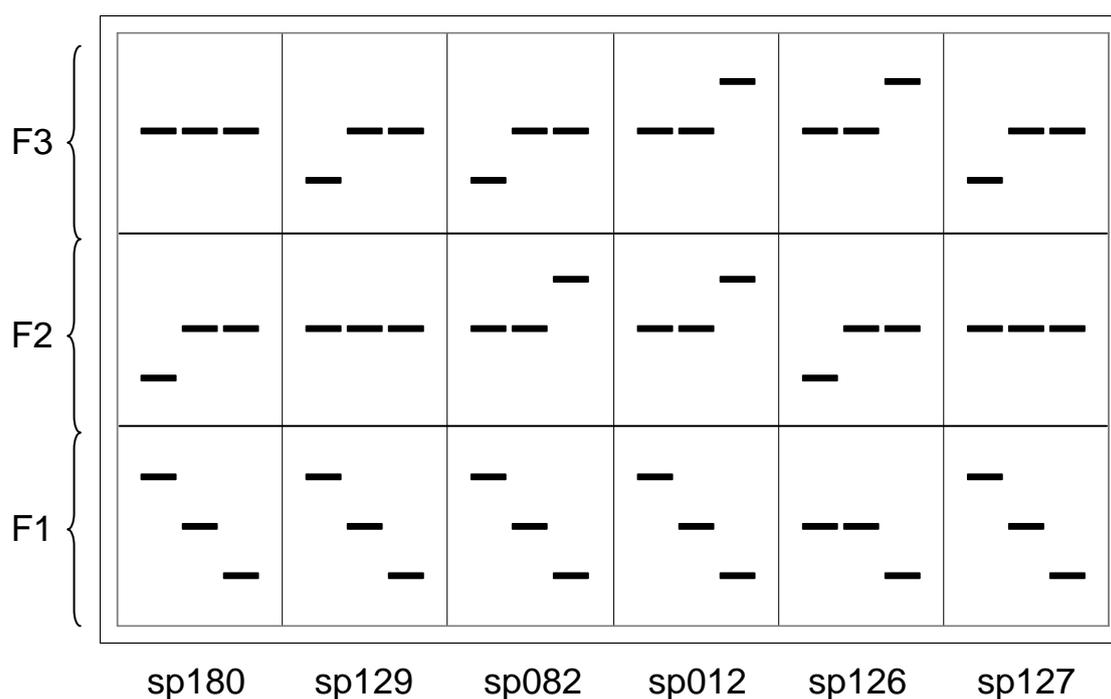


Figure 6.13: Statistically significant changes of F1, F2, and F3 ( $p < 0.05$ ) of the vowel /a/ dependent on stress, sentence reading task. Within each column denoting the speakers, the leftmost crossbar denotes the relative formant frequency position of primary stressed vowels, the middle crossbar the one of secondary stressed vowels and the rightmost crossbar the one of unstressed vowels.

Figure 6.13 and the following figures read as follows. The results of each speaker are presented in a separate column. The main columns are further subdivided, where the leftmost crossbar stands for the relative position of a given formant of primary stressed vowels, the middle crossbar for the relative position of a given formant of secondary stressed vowels, and the rightmost crossbar for the relative position of a given formant of unstressed vowels. The row for each formant is again further subdivided, the highest crossbar denoting a higher formant frequency position relative to the middle and the lowest crossbars, and the lowest crossbar a lower formant frequency position relative to the middle and higher crossbars. The figure gives information about stress-dependent direction of change, but no information about absolute formant frequency values.

Figure 6.13 demonstrates that for the vowel /a/, F1 plays a dominant role in the indication of stress: all speakers, except speaker sp126, expose the highest F1 values in primary stressed vowels and the lowest F1 values in unstressed positions. F1 values of secondary stressed vowels are in between. The results are not as straightforward for F2 and F3. Moreover, speakers deal differently with the way stress is expressed. The speakers sp180, sp129, and sp127 do not differentiate secondary stressed /a/ from unstressed /a/. Speaker sp012 does not differentiate primary stressed from secondary stressed /a/. The speakers sp082 and sp126 differentiate all three stresses with means of F2 and F3. It can be concluded from these data that secondary stress is expressed by all speakers, predominantly by means of changing F1 (i.e. adjusting the degree of lip opening).

In spontaneous speech, three speakers (sp127, sp129, and sp180) do not differentiate primary stressed /a/ from secondary stressed /a/, whilst the others differentiate primary from secondary stress, either by F1 (speakers sp012 and sp127) or F3 (speaker 082). Unstressed /a/ is differentiated by all speakers. Although three speakers still indicate secondary stress, it has to be noted that the discriminatory strength is decreased insofar as only one formant is used for discrimination.

### 6.6.3.2. The vowel /i/

For the vowel /i/, it can be expected that F1 increases with a decrease of stress<sup>141</sup>, whereas F2 and F3 decrease (i.e. the degree of lip aperture and the degree of constriction are increased, and the length of constriction is shortened with decreasing stress). The change in the formant pattern can be schematized as follows:

F1 primary stress < F1 secondary stress < F1 unstressed

F2 primary stress > F2 secondary stress > F2 unstressed

F3 primary stress > F3 secondary stress > F3 unstressed

Figure 6.14 summarizes the results of the t-tests:

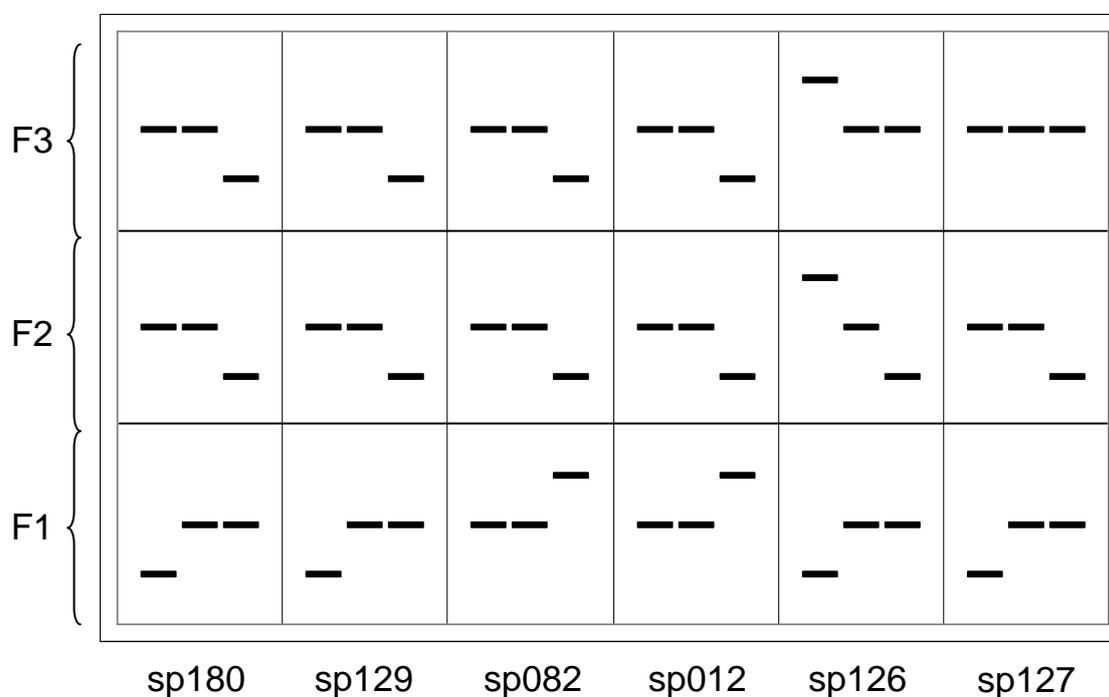


Figure 6.14: Statistically significant changes of F1, F2, and F3 ( $p < 0.05$ ) for the vowel /i/ in dependence on stress, sentence reading task. Within each column denoting the speakers, the leftmost crossbar denotes the relative formant frequency position of primary stressed vowels, the middle crossbar the one of secondary stressed vowels and the rightmost crossbar the one of unstressed vowels.

<sup>141</sup> Several investigations found a lowered jaw position, a greater lip aperture and a more fronted tongue position in stressed vs. unstressed /i/ in English (Harrington et al. 2000, Erickson 2002, Cho 2002). The lowered jaw position and greater lip aperture, which would result in a rise of F1 for stressed /i/, can not be confirmed for Austrian Standard German. See Figure 6.14.

Two speakers (sp082 and sp012) do not differentiate primary from secondary stress. The speakers sp180 and sp129 differentiate primary from secondary stress by F1 and secondary stress from unstressed by F2 and F3. Speaker sp126 differentiates all three stresses by F2, primary stress from secondary stress by F1 and F3, and speaker sp127 differentiates primary stress from secondary stress by F1 and secondary stress from unstressed position by F2. I.e., four speakers indicate secondary stress by changing the formant frequencies of the vowel.

In spontaneous speech, only two speakers realized sufficient items to make a statistical analysis meaningful. Speaker sp126 differentiates primary from secondary stress by F1 (F1 is lower for primary stress) and secondary stress from unstressed position by a higher F2 and a higher F3 for the secondary stressed vowels. Speaker sp127 does not differentiate secondary stressed vowels from unstressed vowels. Again, a number of speaker-specific peculiarities can be observed in how stress is indicated.

### 6.6.3.3. The vowel /ɪ/

In the same way as for the vowel /i/, it can be expected that for the vowel /ɪ/ F1 increases, whereas F2 and F3 decrease as stress decreases. Therefore, the idealized change in formant pattern is the same as for the vowel /i/:

F1 primary stress < F1 secondary stress < F1 unstressed  
 F2 primary stress > F2 secondary stress > F2 unstressed  
 F3 primary stress > F3 secondary stress > F3 unstressed

Figure 6.15 summarizes the results of the t-tests. The results for the vowel /ɪ/ are difficult to interpret. First of all, for those speakers who differentiate stress by means of F1 (sp126, sp127, sp129), the secondary stressed vowels expose a higher F1 than both the primary stressed vowels and the unstressed vowels. In the same way the primary stressed vowels and the unstressed vowels of the speakers sp129, sp012, and sp126 have either a lower (speaker sp129) or a higher (speakers sp012 and sp126) position than the secondary stressed vowels. These four speakers indicate secondary stress, however, in a

way that contradicts the expectations. For the time being, this behaviour cannot be explained. As regards the remaining two speakers, speaker sp082 does not differentiate secondary stressed vowels from the unstressed vowels, and speaker sp180 makes no differentiation at all.

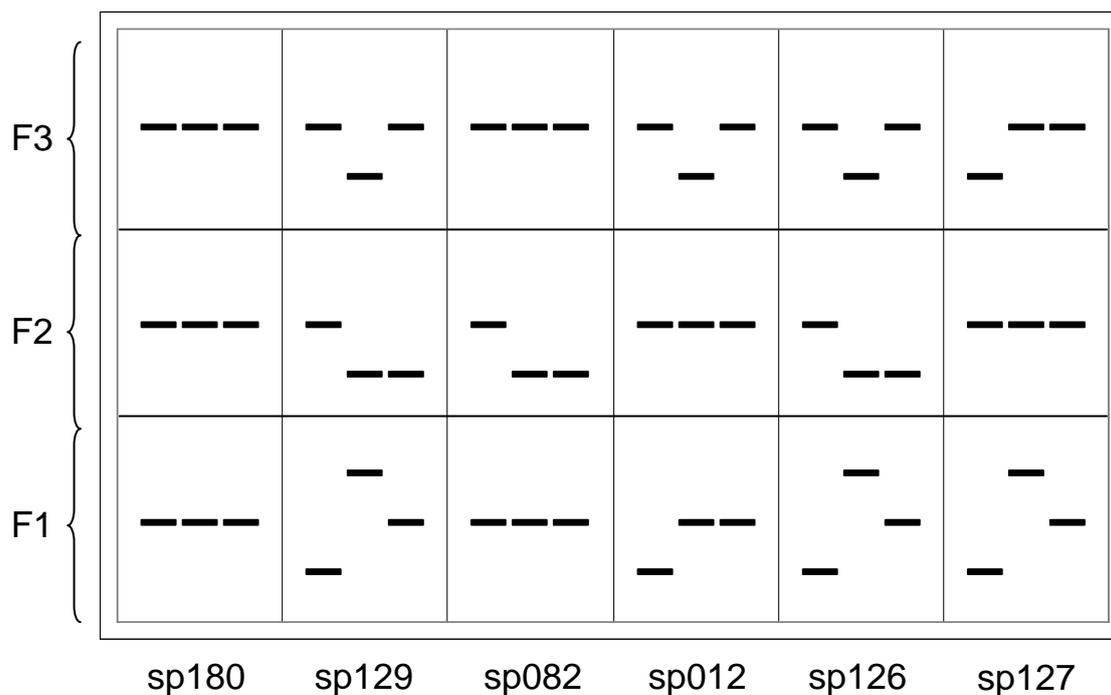


Figure 6.15: Statistically significant changes of F1, F2, and F3 ( $p < 0.05$ ) of the vowel /i/ in dependence on stress, sentence reading task. Within each column denoting the speakers, the leftmost crossbar denotes the relative formant frequency position of primary stressed vowels, the middle crossbar the one of secondary stressed vowels and the rightmost crossbar the one of unstressed vowels.

In spontaneous speech, three speakers produced a sufficient number of items for a statistical analysis. Speaker sp126 did not distinguish any stresses and speaker sp180 did not distinguish secondary stressed and unstressed vowels. Only speaker sp129 distinguished primary and secondary stressed vowels via F2 and F3 and secondary stressed and unstressed vowels via F3.

#### 6.6.3.4. The vowel /ɛ/

It has already been mentioned in 6.6.2 that for the vowels /ɛ/ and /ɔ/, F1 is lowered in the unstressed position as compared to the stressed position. This means that with respect to the degree of lip opening, these two vowels behave in the same way as the vowel /a/, whereas with respect to constriction degree and constriction length, they go with their respective [+ constricted] cognate. The stress-dependent change to be expected should look as follows:

F1 primary stress > F1 secondary stress > F1 unstressed  
 F2 primary stress > F2 secondary stress > F2 unstressed  
 F3 primary stress > F3 secondary stress > F3 unstressed

Figure 6.16 summarizes the results for the sentence reading task:

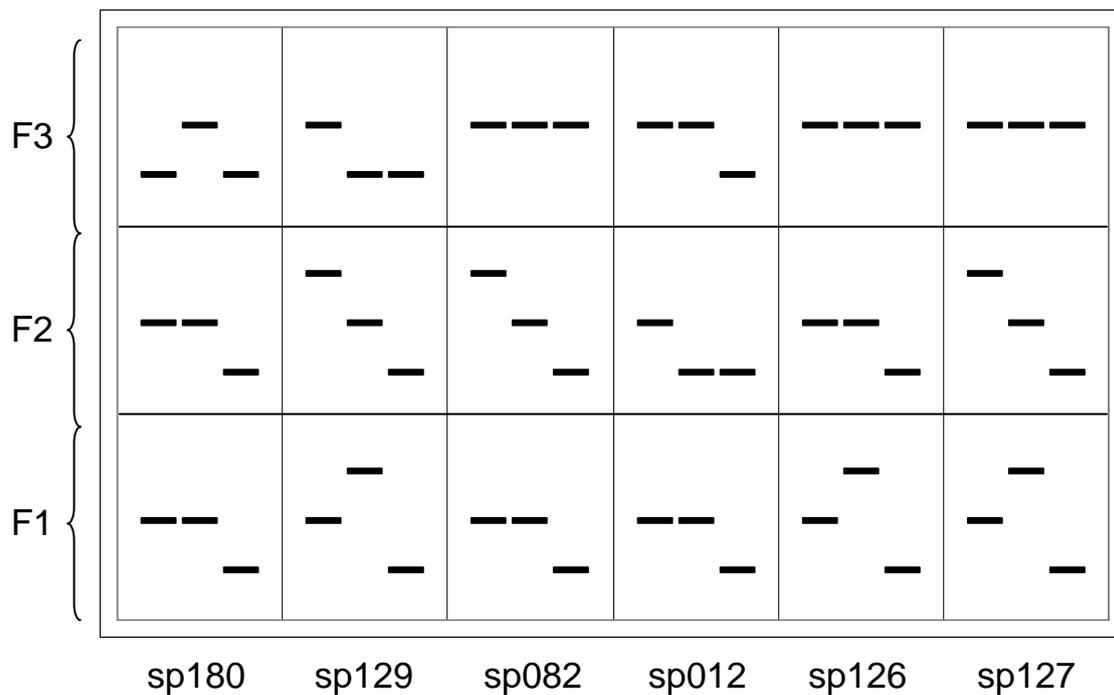


Figure 6.16: Statistically significant changes of F1, F2, and F3 ( $p < 0.05$ ) of the vowel /ɛ/ in dependence on stress, sentence reading task. Within each column denoting the speakers, the leftmost crossbar denotes the relative formant frequency position of primary stressed vowels, the middle crossbar the one of secondary stressed vowels and the rightmost crossbar the one of unstressed vowels.

Figure 6.16 reveals that in every case, primary stressed vowels have a higher F1 than unstressed vowels. However, secondary stressed vowels are not between primary stressed and unstressed vowels. They expose either the same or higher values than the primary stressed vowels. For the time being, no interpretation can be offered for these results.

Secondary stress is either indicated by F2 (by speaker sp129, sp082, sp127), or by combining all three formants. Speaker sp012 distinguishes primary and secondary stress from unstressed vowels via F1 and F2, and primary stress from secondary stress and unstressed vowels via F2. Speaker sp126 distinguishes primary and secondary stress from unstressed vowels by F1. Speaker sp180 distinguishes primary and secondary stress from unstressed vowels in a rather unorthodox way by raising F3 as compared to primary and unstressed vowels, and by distinguishing unstressed vowels from primary and secondary stress by F1 and F2.

In spontaneous speech, again, three speakers produced a sufficient number of items for statistical analysis. The speakers sp126 and sp180 did not distinguish primary from secondary stress. Speaker sp129 distinguished primary from secondary stress via a higher F1 for secondary stressed vowels, and secondary stressed vowels from unstressed vowels via a higher F1 and a higher F2 for the secondary stressed vowels.

#### **6.6.3.5. The vowel /e/**

Sufficient items were only available in the spontaneous speech of two speakers. Speaker sp180 exposed no differences between primary and secondary stress, speaker sp129 distinguished primary from secondary stressed vowels via a higher F1 for the secondary stressed vowels, and the secondary stressed vowels from the unstressed vowels via a higher F2 for the secondary stressed vowels.

### 6.6.3.6. The vowel /u/

Four speakers realized sufficient items in spontaneous speech. Theoretically, F1 and F2 should raise with decreasing stress, whereas F3 should lower, so that the following picture emerges:

F1 primary stress < F1 secondary stress < F1 unstressed  
 F2 primary stress < F2 secondary stress < F2 unstressed  
 F3 primary stress > F3 secondary stress > F3 unstressed

Figure 6.17 presents the results of the statistical analysis for spontaneous speech:

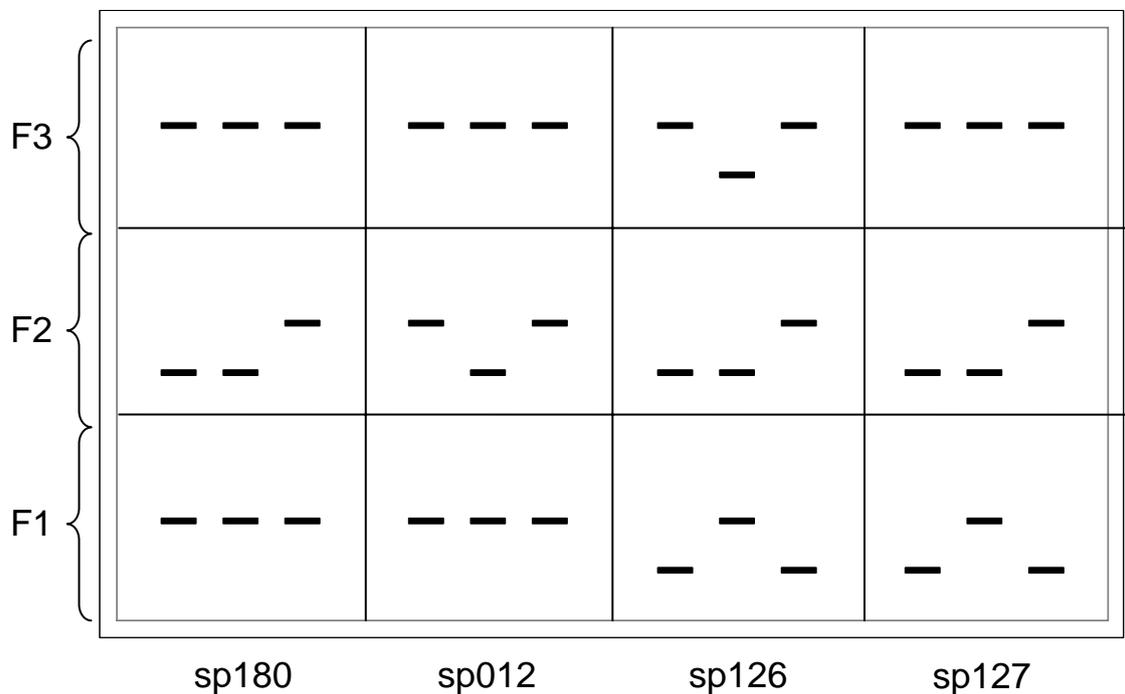


Figure 6.17: Statistically significant changes of F1, F2, and F3 ( $p < 0.05$ ) of the vowel /u/ in dependence on stress, spontaneous speech. Within each column denoting the speakers, the leftmost crossbar denotes the relative formant frequency position of primary stressed vowels, the middle crossbar the one of secondary stressed vowels and the rightmost crossbar the one of unstressed vowels.

Figure 6.17 demonstrates that speaker sp180 does not distinguish primary stress from secondary stress, whereas speaker sp012 does not distinguish primary stress and unstressed positions. Most interestingly, sp012 shows a low F2 only for secondary stress, but not for primary stress. Furthermore, it is noticeable that both speaker sp126

and speaker sp127 mark secondary stress by a higher F1 as compared to both primary stress and unstressed positions, which once more contradicts theoretical assumptions.

#### **6.6.3.7. Summary of the results on secondary stress**

The most salient result is probably the high variability among the speakers<sup>142</sup>. Any speaker has his or her own way to indicate secondary stress by changing the spectral shape of the vowel or by not indicating it at all. Although all speakers seem to agree that primary stress has to be distinguished from unstressed positions, and in what way this should occur, the way to deal with secondary stress does not seem to be codified. However, in spontaneous speech, speakers often reduce the discriminative strength between primary stress and secondary stress on the one hand, and secondary stress and unstressed level on the other hand, down to one formant. It can be concluded, therefore, that a third level exists. Whether and how it is activated, depends on the speaker and on the speaking task. A secondary stress may either be realized in the same way as the primary stress, or in the same way as the unstressed position, or as discrete secondary stress.

#### **6.6.4. Sentence Stress**

The analysis of lexical stress already gives a good insight into the way stress affects the acoustic output of vowels<sup>143</sup>. However, the different relative levels of sentence stress cannot be captured with a sole differentiation of primary, secondary, and unstressed levels (Wagner 2002). Via postlexical stress assignment, the degree of lexical stress can change, providing the respective syllable with either higher or lower prominence. Moreover, in a string of unstressed syllables, not all syllables may be equally

---

<sup>142</sup> High variability among speakers has also been observed by Kleber & Klippahn (2006).

<sup>143</sup> The focus is on vowels. This does not of course mean that consonants are not affected by stress. Jessen et al. (1995) showed that the closure phase of plosives is longer in a stressed syllable. The same results were obtained by Cho & McQueen (2005) for Dutch, Heldner & Strangert (2001) for Swedish, Greenberg et al. (2003) for English.

unstressed. E.g. in a polysyllabic word a lexically stressed syllable can still retain some degree of stress when unaccented, via postlexical stress assignment (Barry & Andreeva 2001). Post-focal words receive lower overall intensity than pre-focal words, both as regards the stressed vowel and the unstressed syllable (Heldner 2003: 50). Additionally, the distance of a word from the focus plays a role: a word directly preceding the focal word is less stressed<sup>144</sup> than a word with a two-word distance (Heldner 2003). Much work has been done on the effect of prosodic boundaries (Dilley et al. 1996 on glottalization in English, Fougeron & Keating 1997 on /n/ in English, Fougeron 2001 on five consonants and two vowels in French, Cho & Keating 2001 on alveolar consonants in Korean, Cho 2004 on /a/ and /i/ in English, Tabain 2003 on /aC/ sequences in French, Tabain et al. 2004 on /uC/ sequences in French, Tabain & Perrier 2005 on /iC/ sequences in French, to name just a few). There is strong agreement that different prosodic boundaries result in different outputs. However, prosodic boundaries exercise less effect on /i/ than on /a/ in French. The prosodic boundary accounted for 43% of the variability of /a/, but only for 13% of the variability of /i/ (Tabain & Perrier 2005)<sup>145</sup>. As concerns the vowel /a/, F1 increases and F2 decreases with increasing prosodic strength of the boundary (Tabain 2003). For /u/, both F1 and F2 decrease (Tabain et al. 2004), and for /i/, the length of the front cavity seems to be primarily affected, resulting in variability of F3 (Tabain & Perrier 2005).

According to Möhler & Dogil (1995) and Féry & Herbst 2004, the primary acoustic correlate of sentence stress (prominence) is fundamental frequency. In their analysis of seven dialects of British and Irish English, Kochanski et al. (2005) found loudness, together with duration, the most robust indicators marking prominent syllables in a sentence, whereas fundamental frequency contributed little. However, the contribution of the spectral change of a vowel (or further segments) to perceived promi-

---

<sup>144</sup> Wagner (2002) mentions “präfokale Deakzentuierung” as one strategy for assigning prominence to the following syllable.

<sup>145</sup> /i/ in French is said to be stable (Piternann 2000, Tabain & Perrier 2005).

nence should not be neglected either. Fant et al. (2000b), in their analysis of prominence in Swedish, name six parameters responsible for prominence in the following order of relevance: duration, fundamental frequency, overall intensity, overall intensity with a high-frequency pre-emphasis, voice source, and spectral modifications of consonants and vowels. They found out that a rise of F1 and a lowering of F2 lead to higher perceived prominence for the vowel /a/. The high vowels, on the other hand, attained a higher  $R_s$ <sup>146</sup> by tightening the degree of constriction. How, and to what extent, relative prominence is realized in actual speech depends on the individual speakers (Fant et al. 2002). I.e. speaker-specific differences are to be expected in the realization of prominence, as exemplified in Figure 6.18.

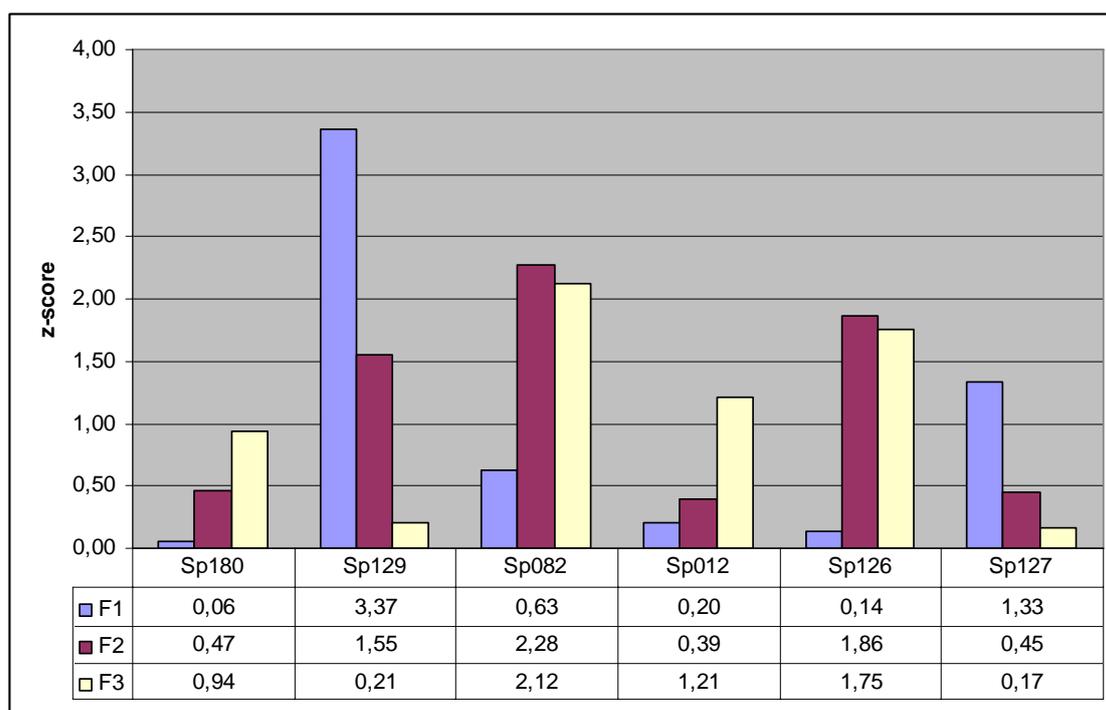


Figure 6.18: Z-deviation F1, F2, and F3 of the vowel /i/ from “liebe” (dear: ADJ) from the vowel /i/ from “Liebe” (love: N), sentence reading task. For each speaker, n = 4. With  $z \geq 1.5$  no correspondence of the respective formant is given any more. The values have been root squared to facilitate reading. Sp = speaker.

<sup>146</sup>  $R_s$  is a continuous interval scale from 0 to 30 indicating perceived prominence. In perception tests conducted by Fant et al. 2000a, content words received an average  $R_s$  of 18.6, function words an average  $R_s$  of 11.

Figure 6.18 gives the deviation for each formant of the vowel /i/ from “liebe” (dear: ADJ) from the vowel /i/ from “Liebe” (love: N)<sup>147</sup>. Though all /i/s carry primary lexical stress, the vowel /i/ of the adjective is supposed to bear less stress than the vowel /i/ of the noun. As concerns spectral modifications of the vowel /i/, three speakers (sp180, sp012, and sp127) make no difference between the noun and the adjective ( $z < 1.5$ ). Speaker sp129 has quite a substantial difference in F1, with lower F1 values for the noun, indicating a greater mouth opening in the adjective, and a difference in F2, with a higher F2 value for the noun, i.e. a tighter constriction. Both speakers sp082 and sp126 show differences for F2 and F3, both being higher in the noun, as expected, indicating a tighter and longer constriction of the tongue with a slightly more forward tongue position in the noun.

In the same way as a specific part of speech, the role in the sentence might show its effect on the spectral shape of a given vowel. In Figure 6.19, the effects of a preceding focal word on the stressed syllable of “vergessen” (to forget)<sup>148</sup> are presented. Only one speaker (sp180) shows a difference between the two positions in the sentence, insofar as the post-focal position triggers a lower value for F3. For all the other speakers, the lexically stressed vowel /ɛ/ of “vergessen” is realized in the same way, regardless of sentence position. Pre-focal distance seems to be of higher relevance; Figure 6.20 shows the deviation of the vowel /i/ from “diesen” (these: DEM: DAT: PL), which holds the second position before the noun, from the vowel /i/ from “dieser” (this: DEM: NOM: SG)<sup>149</sup> which stands directly before the noun.

<sup>147</sup> The context of the items was: “unsere liebe Ruth” (our dear: ADJ Ruth) and “die Göttin der Liebe” (the goddess of love: N).

<sup>148</sup> The contexts were: “jetzt hat sie doch glatt vergessen” (now she’s plain forgotten, postfocal position) and “Leider hab ich vergessen“ (unfortunately I’ve forgotten, default stress position).

<sup>149</sup> The contexts were: “Bei diesen wilden Jagden” (at these: DEM: DAT: PL wild huntings) and “..., daß dieser Kopf ein ausgezeichnetes Modell [...] ist” (... that this: DEM: NOM: SG head is an excellent model).

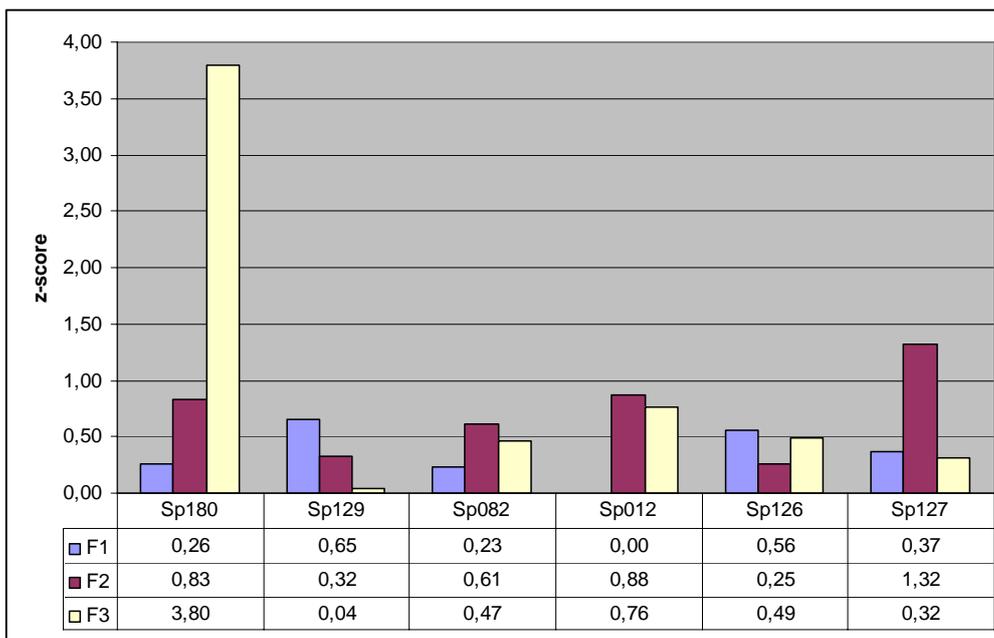


Figure 6.19: Z-deviation F1, F2, and F3 of the vowel /ε/ from “vergessen” (to forget) in postfocal position from the vowel /ε/ from “vergessen” in a default stress position, sentence reading task. For each speaker, n = 4. With  $z \geq 1.5$  no correspondence of the respective formant is given any more. The values have been root squared to facilitate reading. Sp = speaker.

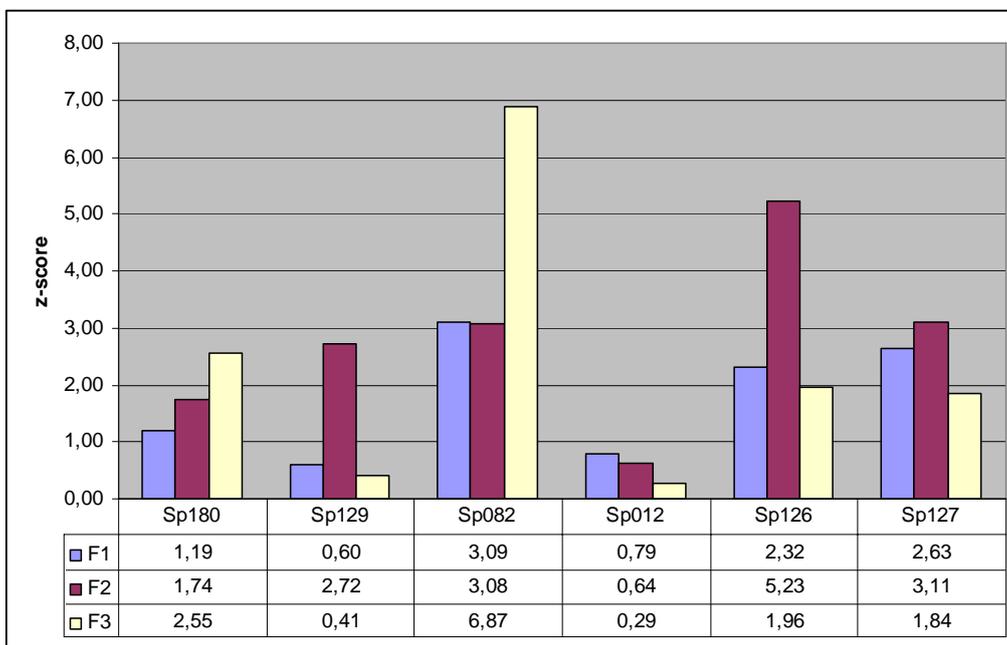


Figure 6.20: Z-deviation F1, F2, and F3 of the vowel /i/ from “diesen” (these: DEM: DAT: PL) holding the second position before the noun from the vowel /i/ from “dieser” (this: DEM: NOM: SG) standing directly before the noun, sentence reading task. For each speaker, n = 4. Explanation as in Figure 6.19.

In this case, only one speaker (sp012) makes no difference between the two positions. All the other speakers differentiate either one formant (sp129), two formants (sp180), or all three formants (sp082, sp126, sp127). The direction of the change is in accordance with the results on intensity presented in Heldner (2003): the vowel of /dis/ directly preceding the noun shows a higher F1 and a lower F2 and F3. From an articulatory point of view, the degree of mouth opening and the degree of constriction (tongue-palate distance) are enlarged, whilst the length of constriction is possibly shortened.

### **6.6.5 The rhythm of speech**

The above examples vividly show that a) relative sentence stress affects the spectral shape of the vowels and b) the realization of relative sentence stress is highly speaker dependent. The reason for the existence of such graded differences lies in the inherent rhythmic behaviour of speech which requires a succession of foregrounding and backgrounding. Sentence stress finds its expression only in relation to the other entities in a sentence or an utterance. Therefore, it is strongly connected to the rhythm of speech or, to put it more explicitly, the relative stress levels assigned to phonemes, syllables, and words, which make up what is perceived as a rhythmic behaviour in the so-called stress-timed languages.

Rhythm is generally highly connected with timing; with ideal beat intervals of about 300 ms to 400 ms (Schreuder 2006: 101). In so-called stress-timed languages, these intervals are supposed to stay equally long regardless of the amount of unstressed syllables in between. This assumption has triggered a further, implicit hypothesis, namely that the degree of reduction imposed on unstressed syllables depends on the amount of unstressed syllables between two stresses. The isochrony hypothesis was never confirmed, leading to other – statistical – methods to classify languages (see e.g. Low & Grabe 1995, Low et al. 2000, Grabe & Low 2002, Ramus et al. 1999, Ramus 2002, Galves et al. 2002, Rouas et al. 2005, see Cummins 2002, Wagner 2002 for a

critique). Ramus et al. (1999) assume that stressed-timed languages allow more complex syllables than syllable-timed languages and classify languages according to the percentage of duration taken up by vocalic intervals (%V) and the standard deviation of the duration of consonantal intervals within a sentence ( $\Delta C$ ). The approach of Low & Grabe (1995) and Grabe & Low (2002) “took a direct route from impressionistic observations of rhythmic differences between languages to the acoustic signal” (Grabe & Low 2002: 519). The rhythmic differences should be reflected in the duration of vowels and the duration of intervals between vowels. Additionally, Low et al. (2000) analysed the degree of reduction of vowels (F1, F2), which they found to be higher in British English than in Singapore English. Galves et al. (2002), on the other hand, assume that rhythmic class discrimination is based on a rough measure of sonority. Rouas et al. (2005) relied on pseudo-syllabic patterns by automatically segmenting the speech chain in vowel and non-vowel segments. The results of these approaches more or less corroborated the traditional distinction of languages as stress-timed, syllable-timed, and morae-timed. The analyses discussed above are mainly based on diverse durational relations, without questioning whether duration is really the basic unit that typologically keeps languages apart<sup>150</sup>. However, what is even more important is the fact that rhythm is conceived as an alternating dichotomic pattern of “a” and “b”, where “a” and “b” can, in principle, stand for anything: an alternating succession of vowels and consonants or consonant clusters, of sonority and non-sonority, of stressed and one or more unstressed syllables, etc.

Figure 6.21 gives an example where speech is not feasible in a mere dichotomous a – b pattern, but might even be asynchronous to rhythms performed by the body. During the recording of Albanian spontaneous speech, the interviewer started to rhythmically beating with his finger on the table.

---

<sup>150</sup> Rouas et al. point out that “rhythm cannot be reduced to a raw temporal sequence of consonants and vowels” (2005: 453).

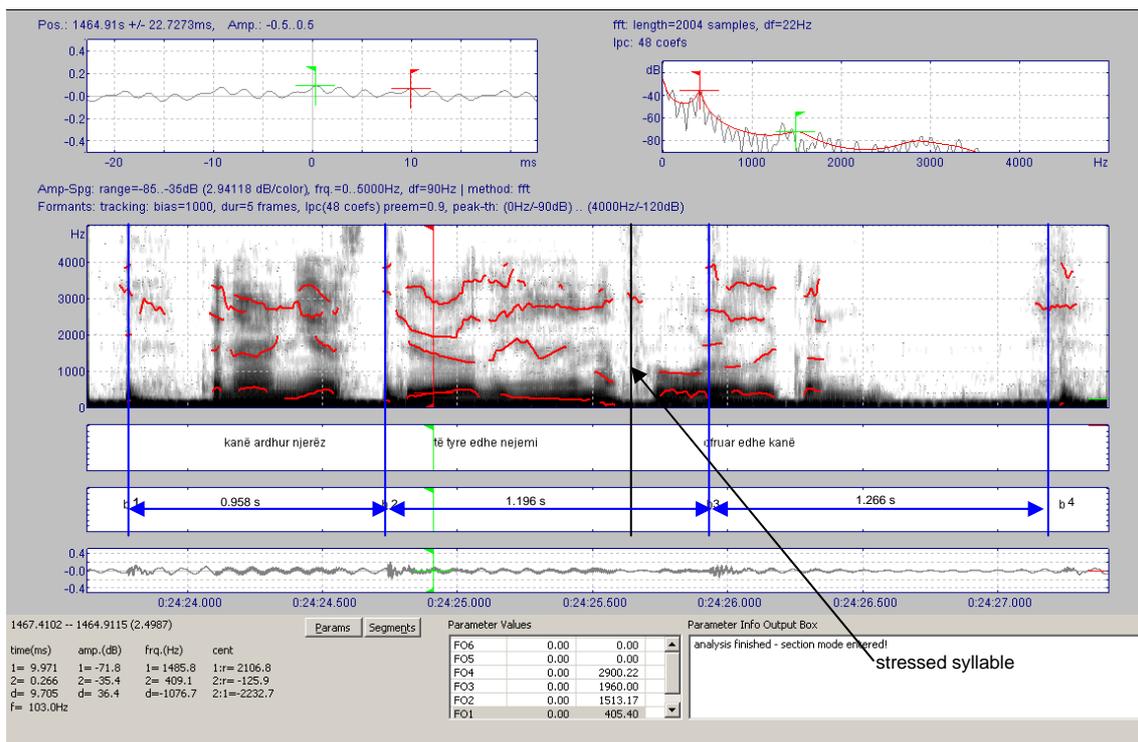


Figure 6.21: Spectrogram of the sequence “... kanë ardhur njerëz të tyre edhe ne jemi afruar edhe kanë...” (...their people came and we went there and they have...) Language: Albanian, male speaker. Bottom panel: waveform window, next panel from bottom: duration of the single beats, 3<sup>rd</sup> panel from bottom: transcription, 4<sup>th</sup> panel from bottom: spectrogram window, left upper panel: waveform zoom window, right upper panel: amplitude spectrum window.

In the spectrogram of Figure 6.21, these beats are marked with a blue vertical line, the duration between the beats is shown in the second panel from the bottom. The duration between each beat is approximately 1 s. The second beat is synchronous with the plosive release of the stressed syllable “tyre”. The third beat, however, starts later than the next stressed syllable “ofruar” (primary stress on the second syllable “fruar”), indicated by the black line in the spectrogram. The duration from the beginning of the stressed syllable “tyre” to the beginning of the next stressed syllable “fruar” is 0.899 s, therefore, only 60 ms are missing to synchronise the third beat with the beginning of the stressed syllable and to equalise the duration of beat 1 to beat 2 with beat 2 to beat 3. In order to achieve this, the speaker would have had to dispense with some backgrounding processes in the preceding unstressed syllables “edhe ne jemi”. Instead of reducing

“nejemi” to one syllable [nēm], articulating all three syllables would have easily compensated for the missing 60 ms. Providing an unstressed sequence with the same articulatory accuracy is, however, in contrast to the principle of backgrounding and foregrounding. Therefore, an exact timing is of secondary relevance. The primary concern in the rhythm of speech is to properly background unstressed sequences in order to contrast them with the stressed positions. How much time is required for the accomplishment of this activity, depends on the content that has to be conveyed. And the syntax and semantics of a language do not necessarily consider that a required amount of unstressed syllables is available in order to fit an exact timing, or, to put it another way, a speaker of a language plans the sequences of words he or she wants to utter and the processes that have to be applied, but the speaker does not search for another word with, e.g., more consonant clusters in order to arrive at an exact timing.

Therefore, speaking is different from walking, dancing, playing an instrument, or even poetic rhythm, in that it is to a larger extent a social activity. Speakers want or have to convey in some way or other more and less important units<sup>151</sup> of speech, and they convey these units in a graded way of foregrounding or backgrounding them. In an utterance, not only important and non-important information exists, but also further levels of importance, new information, redundant information, repetitions, etc. Therefore, in an analysis of the rhythm of speech, dichotomous conceptualizations have to be given up. Moreover, it has to be questioned whether speaking is really ruled by timing or whether there is some sort of interaction between what a speaker wants to say, the speech situation<sup>152</sup> he or she is in, and the time he or she computes or has at his or her disposal for the planned utterance. Therefore, the gradings of foregroundings and backgroundings might differ considerably in dependence of diverse external factors.

---

<sup>151</sup> Unit has a broad meaning here and can be new or old information, a phrase, a word, a syllable, or a phoneme.

<sup>152</sup> For example, in an oral exam situation, a speaker might exhibit a different rhythmic patterning than in an oral presentation, although the amount of time they have at their disposal might be the same.

Therefore, rhythmic patterning is not only achieved by an alternating succession of foregrounding and backgrounding certain phonemes, syllables, or words, but by additionally grading the amount of foregrounding or backgrounding. This principle of figure and ground is well embedded in the semiotic model elaborated by Dressler (1980, 1985, 1996, 2002) and has been exemplified by Madelska & Dressler (1996) on Polish and Czech. The graded foregroundings and backgroundings can be expressed in various ways, whereby the speaker can make use of the whole speech production apparatus. This includes both prosodic and segmental features, which, from a production point of view, cannot be clearly separated (Fant & Kruckenberg 2004: 249). The interplay of parameters is impressively shown in the multi-parameter analysis of Swedish prosody performed by Fant & Kruckenberg (2004). Which parameters are enhanced and which play a secondary role, or whether all parameters are relevant, is language-specific and speaker-specific.

(Standard Austrian) German is a language which assigns different levels of stress to units with different importance. It has already been outlined in the previous subchapters that the production of vowels is affected by both pre-lexical and post-lexical stress assignment. However, unlike Russian or Bulgarian, German shows no stress-dependent vowel-quality alternation (Barry & Andreeva 2001). Therefore, with respect to vowels, the alternating succession of graded foregroundings and backgroundings is expressed by modifying the degree of constriction, the length of constriction, the degree of lip protrusion, and the degree of lip aperture, whilst the location of the constriction stays the same. Again, which parameters are preferred depends on vowel-type and on the speaker as well.

The interaction of these graded foregroundings and backgroundings and of rhythm is exemplified for the vowel /a/<sup>153</sup> in the following sentence:

---

<sup>153</sup> The vowel /a/ has been chosen, because, unlike the other vowels, it is not paired and therefore more tokens for comparison can be found in a sentence.

Sentence 065:  
 Das Paar hat-t-e sich in Leipe, einer  
 DEF:N:SG AUX couple AUX-PST-3S REFL:3S ILL Leipe INDEF-F:SG:DAT  
 klein-en Stadt in Deutschland, nieder-ge-lass-en.  
 small-F:SG:DAT town ILL Germany VPREF-PP-settle-PP

In principle, the sentence is designed in an iambic rhythm which demands a weak node on the first syllable of “hatte”, which is nevertheless supposed to be stronger than the definite article “das”, due to lexical primary stress on the first syllable, and a weak node on the second syllable of “Deutschland”, which, as a compositum, might nevertheless retain some degree of stress. “Stadt”, being the head of the adposition, receives less prominence than “Paar”. The fourth syllable of “niedergelassen” has a secondary stress, and it can be assumed that the vowel is slightly stronger than the one in “Deutschland”. Therefore, as concerns the vowel /a/, the hierarchy of stresses should be as follows:

**Paar > Stadt > niedergelassen > Deutschland > hatte > das**

With respect to the modification of spectral shape, F1 is supposed to gradually decrease with decreasing prosodic strength, F2 and F3 are supposed to gradually rise.

As regards F1, the vowel /a/ in “Paar” received the highest value by all speakers. Five speakers (sp180, so129, sp082, sp012, and sp126) attributed a lower or at least equal<sup>154</sup> F2 value to the vowel /a/ in “Paar”. Sp127 had a lower F2 in “Deutschland”. Four speakers (sp180, sp129, sp126, sp127) exposed a lower or equal F3 in “Paar”. Sp082 attributed lower F3 values to the vowels /a/ of “Stadt” and “niedergelassen”. Sp012 had a lower F3 value in “Stadt” and “hatte”. In sum, every person attributed a higher F1 and either a lower F2 or a lower F3 to the vowel /a/ in “Paar”. Therefore, this vowel received highest prominence by all speakers. This result corroborates the observations made by Wagner (2002) that speakers show a high agreement both on the placing and the modelling of primary stress. Moreover, as regards the vowel /a/, F1 seems to be of higher relevance for modelling stress than either F2 or F3.

---

<sup>154</sup> Means in this case that the negative z-score (= F2 is lower in another vowel /a/ than in the vowel /a/ of “Paar”)  $\leq 1.5$ .

Since the vowel /a/ in “Paar” received highest prominence by all speakers, it appears justified to look at the deviation of the formant frequencies of the other vowels from the vowel /a/ in “Paar”. Figure 6.22 presents the results on F1:

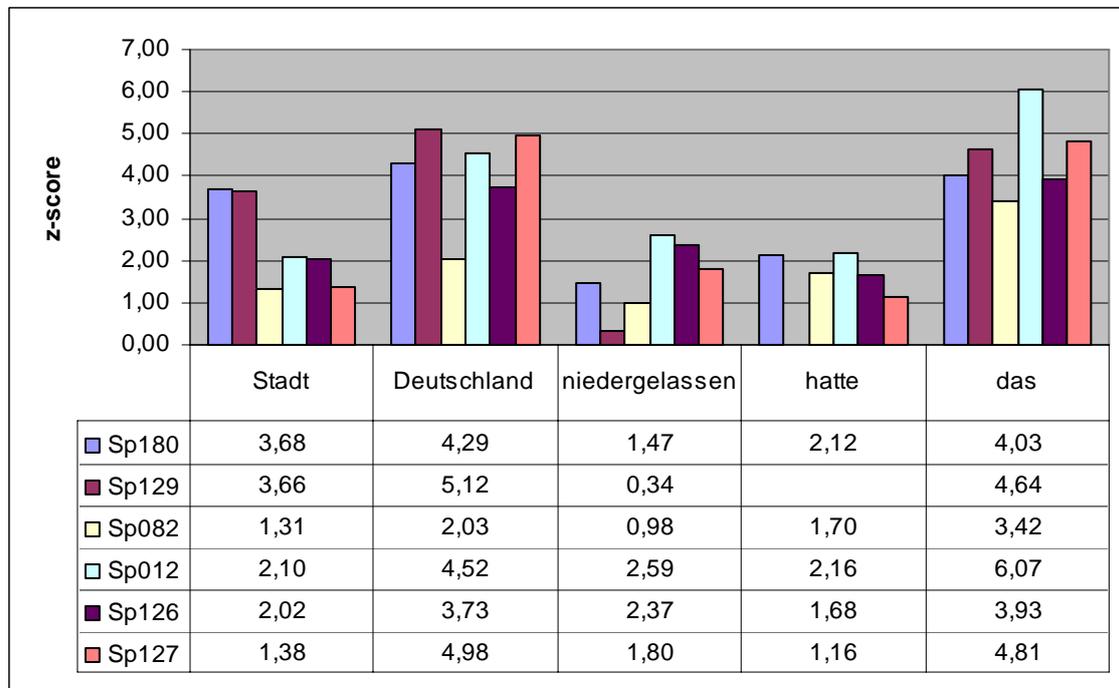


Figure 6.22: Z-score of F1 of the vowels /a/ from the vowel /a/ in “Paar” (couple). With  $z \geq 1.5$  no correspondence of the respective formant is given any more. Sp = speaker, sentence reading task.

According to Figure 6.22, all speakers confirm in “das” deviating most and substantially from “Paar”, i.e. F1 is substantially lower in “das”. For the rest of the vowels, speaker-specific differences can be observed. The /a/ in “Deutschland” receives a very low F1 as well, and is, in principle, as unstressed as “das”. Only speaker sp082 and speaker sp012 differentiate between “das” and “Deutschland”. Most interestingly, F1 of “hatte” deviates to a much lesser degree than expected. Four speakers (sp082, sp012, sp126, sp127) assign approximately the same degree of lip aperture to the vowel /a/ in “hatte” as to the vowel /a/ in “Stadt”, even though the two vowels in “Paar” and in “hatte” are not separated by an unstressed syllable. However, the relatively higher F1 in “hatte” is not to be interpreted as a provoked stress clash, but rather as an assimilation; since it is

not necessary to change the degree of lip aperture during the /h/, lip aperture is simply not modified in the vowel of “hatte”, and the decrease of stress is modelled via F2 and F3 by all speakers except speaker sp012 (see Figure 6.23).

Three speakers (sp180, sp129, sp082) attribute the same degree of lip aperture as in “Paar” to the vowel /a/ in “niedergelassen”. The other three lower F1 to about the same degree as in “hatte”.

“Stadt” exerts the most salient speaker specific differences: Two speakers (sp180 and sp129) lower F1 substantially, even more than in “hatte” and “niedergelassen”. Two further speakers (sp012 and sp126) lower F1 to about the same degree as in “niedergelassen” and “hatte”, and two (sp082 and sp127) do not differentiate “Stadt” from “Paar”.

Regarding the z-scores of F2 and F3<sup>155</sup>, which seem to work together, the results on F1 are either enhanced or stay the same, but they are never reduced (see Figure 6.23).

The low F1 values of “das” are further enhanced by speaker sp180, sp129, sp082, sp126, and sp127, who expose a rise of F2 and/or F3. Only speaker sp012 does not further set off the unstressed position via F2 and F3; his F1 values are the lowest of all speakers anyway. A further contribution of F2 and F3 does not seem to be necessary. “Deutschland”, however, is only enhanced by speaker sp180 and sp082, so that, on the whole, “Deutschland” is slightly more stressed than “das”. It has already been stated that for the vowel /a/ in “hatte”, less deviation from the primary stressed position has been observed with respect to F1 than hypothesized. This assimilation is balanced by a higher deviation of F2 and F3 for all speakers except speaker sp012, and especially for speaker sp082.

---

<sup>155</sup> The mean of the z-score of F2 and the z-score of F3 has been calculated.

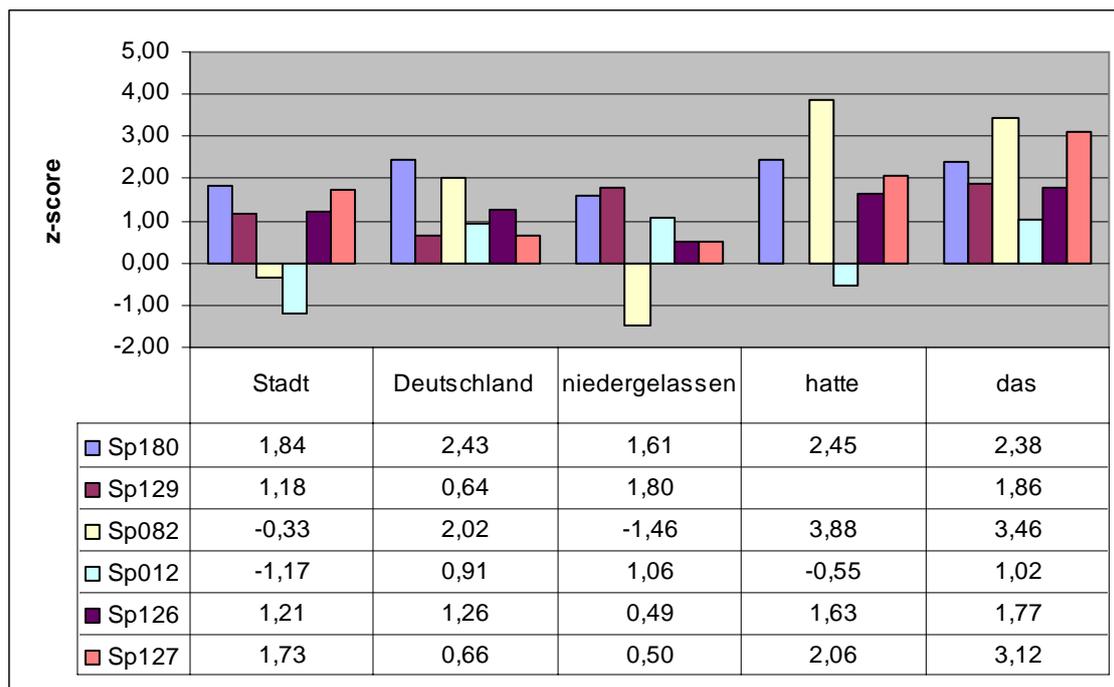


Figure 6.23: Mean z-deviation of F2 and F3 of the vowels /a/ from the vowel /a/ in “Paar” (couple). With  $z \geq \pm 1.5$  no correspondence of the respective formant is given any more. Sp = speaker, sentence reading task. A negative value indicates an F2 or F3 value contrary to the hypothesis.

With respect to F1, the three male speakers (sp012, sp126, sp127) show a decrease in “niedergelassen”, whereas the three female speakers (sp082, sp129, sp180) do not differentiate “niedergelassen” from “Paar”. However, as regards F2 and F3, two female speakers (sp180 and sp129) raise F2 and F3 in comparison to “Paar”. Speaker sp082, however, treats “niedergelassen” in the same way as “Paar”, i.e. providing it with a primary stress. The explanation for the relative prosodic strength of “niedergelassen” can be found in its preceding a strong boundary, which triggers a higher F1 and a lower F2 for the vowel /a/ (Tabain 2003).

“Stadt” is only further enhanced by speaker sp180. All the others do not differentiate “Stadt” from “Paar” with respect to F2 and F3. Therefore, “Stadt” is only differentiated via F1.

It can be concluded that speakers agree on the most stressed and the most unstressed position and model them in the same way (in the case of the vowel /a/ pre-

dominantly via F1). The rankings of the others show speaker-specific differences which can be summarized as follows:

Speaker sp180: Paar > niedergelassen > Stadt, hatte > Deutschland, das  
 Speaker sp129: Paar > niedergelassen > Stadt > Deutschland > das  
 Speaker sp082: Paar, niedergelassen, Stadt > Deutschland > hatte > das  
 Speaker sp012: Paar > niedergelassen, Stadt, hatte > Deutschland > das  
 Speaker sp126: Paar > niedergelassen, Stadt > hatte > Deutschland > das  
 Speaker sp127: Paar > niedergelassen, hatte, Stadt > Deutschland > das

Except for the most stressed and the most unstressed position, these rankings differ from the initial hypothesis. The gradings of foregroundings and backgroundings play an important role and exert a substantial influence on the spectral shape of vowels as well. Figure 6.24 gives an overview of how the rhythmic patterning is performed by the speakers:

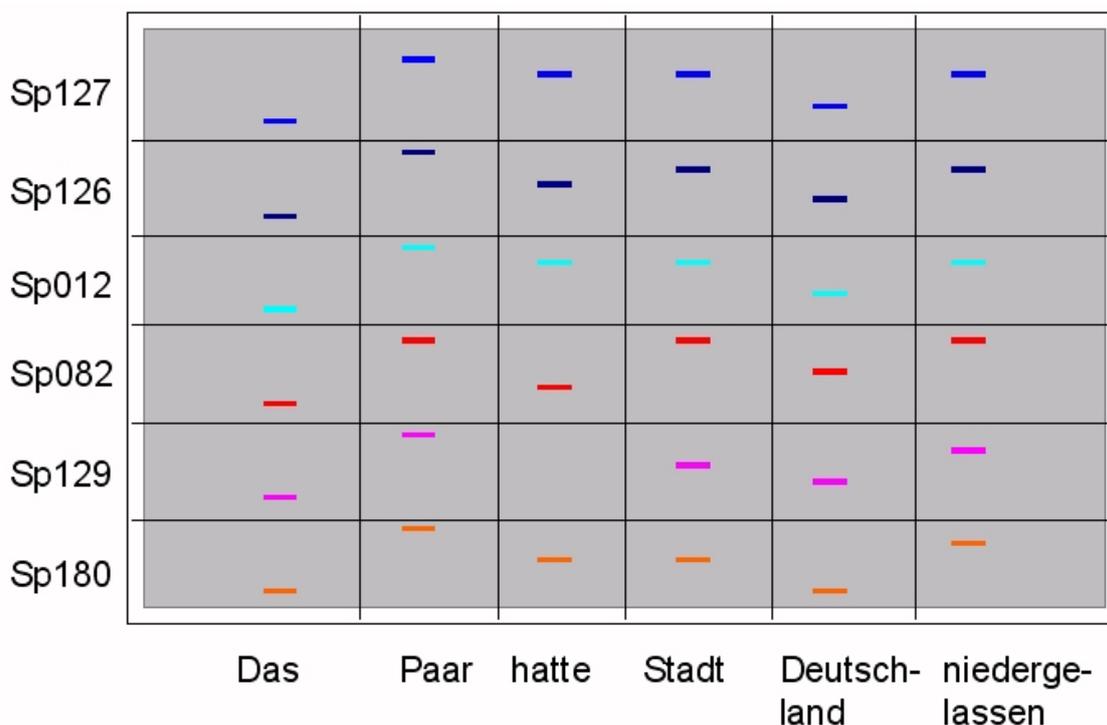


Figure 6.24: Changes of formant frequencies of the vowel /a/ in dependence of stress-conditioned foregrounding and backgrounding. The highest crossbar is most foregrounded, the lowest crossbar is least foregrounded. Each speaker is assigned a different colour.

Figure 6.24 demonstrates that the respective spectral changes of the vowels evolve into a rhythmic pattern. However, there are some degrees of freedom with respect to how

and to what extent the various stresses are shaped differently from the most stressed and the least stressed ones. These degrees of freedom are exploited by the speakers and might be additionally responsible for the fact that speakers are perceived differently<sup>156</sup>. Therefore, if speaker-specific differences are to be found, low-level analyses have to be performed. An overall comparison of e.g. certain vowels blurs most of the differences, as will be shown in the next chapter.

---

<sup>156</sup> Widera & Portele (1999) and Widera (2002) report that listeners distinguish three to five reduction levels depending on the vowel. However, listeners are able to compensate for speaker-specific differences (Widera 2002: 191).

## 7. Speaker-specific traits

Throughout the analysis on vowels in Standard Austrian German, speaker-specific traits emerged. These especially showed up in the dealing with variation, but speaker-specific traits also occurred in the phonemic representation ([±constricted] vs. [±open]). Therefore, if one wants to look at speaker-specific traits, one has to go into detail. Approaches which pool together all vowels (Grigoras & Nolan 2005)<sup>157</sup>, run the risk of blurring most of the speaker-specific information and might arrive at correspondences where there are none. For example, one-way ANOVA calculated over all stressed vowels of spontaneous speech renders no differences between the three male speakers for F2 ( $F_{2,371} = 0.33$ ,  $p = 0.72$ ). As regards F3, differences show up for speaker sp127, but none between sp012 and sp126. F1 renders no differences between speaker sp012 and sp127, but differences for speaker sp126. Given the fact that in true forensic cases, F1 can hardly ever be mobilised because of the frequency band limitation of the telephone (see Künzel 2001 and Nolan 2002 on the usability of formants in speaker identification), an analysis of the three male speakers would result in a correspondence of speaker sp012 and speaker126. As concerns the female speakers, all three speakers differ for F1, but, again, two speakers (sp129 and sp082) cannot be kept apart with respect to F2 and F3 in spontaneous speech.

The situation is slightly better for the unstressed vowels of spontaneous speech. All three male speakers are kept apart via F3. As concerns F2, speakers sp126 and sp127 are differentiated, whilst speaker sp012 is not. The three female speakers are kept apart by F1, F2, and F3. Therefore, as has been pointed out in Moosmüller (2002), based on the theoretical considerations put forward in Dressler (1979), in forensic

---

<sup>157</sup> Grigoras & Nolan (2005) point out themselves that such an analysis has to be extended by an additional detailed analysis.

casework, unstressed positions are more reliable than stressed ones. Nevertheless, given the small sample of only six speakers which already renders correspondences where there are none, one should apply such methods with extreme caution.

Proceeding to particular vowels in spontaneous speech, the speakers cannot be differentiated either. Figure 7.1 shows the density plot of all stressed /a/ vowels of spontaneous speech.

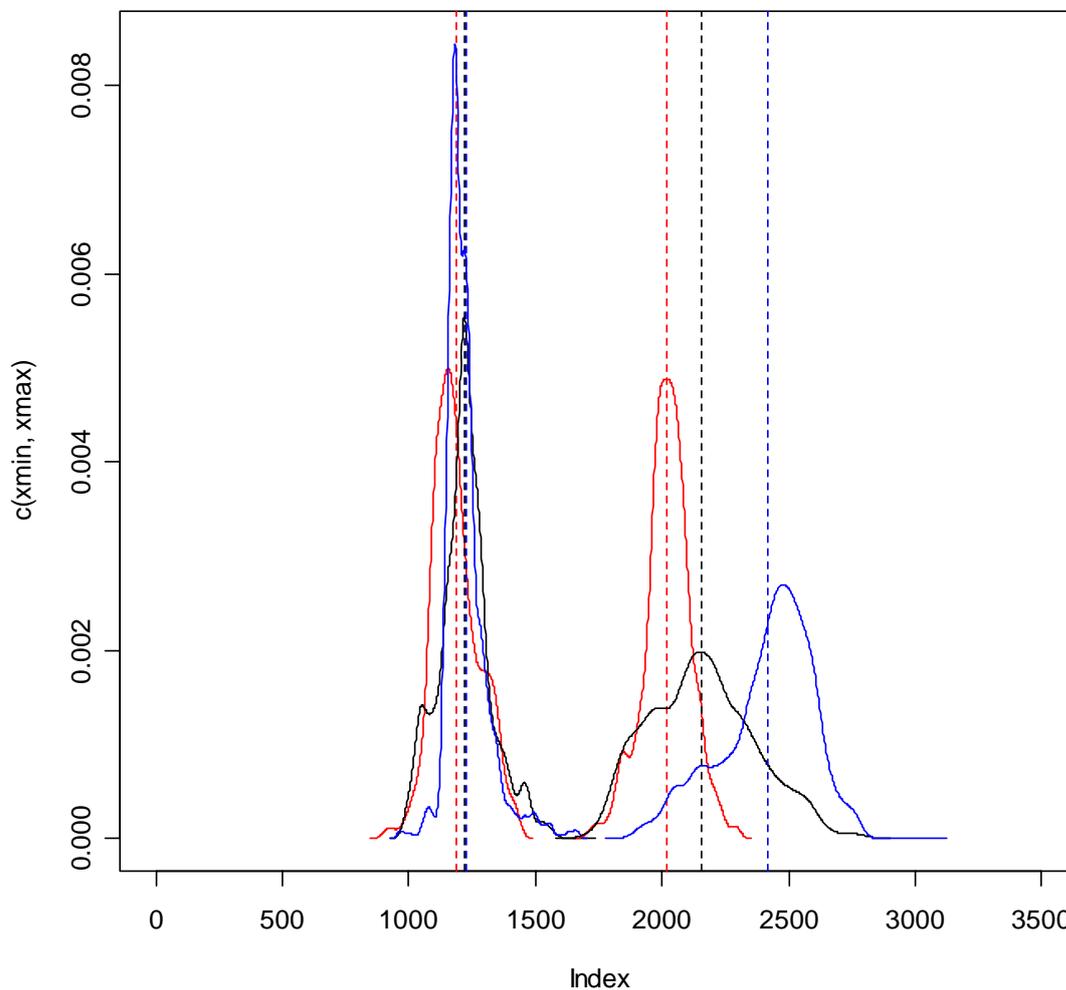


Figure 7.1: Density plot of F2 and F3 of the vowel /a/ in stressed position in spontaneous speech. Red: speaker sp012, black: speaker sp126, blue: speaker sp127.

From Figure 7.1 it can be easily read that the three speakers are differentiated by F3, but not by F2. Table 7.1. to 7.9 summarize the results of the two-tailed t-tests for all vowels<sup>158</sup> in stressed and unstressed position in spontaneous speech of the six speakers. For better readability, the “+” are highlighted. F1 is added for the sake of completeness.

/a/	Sp012 – Sp126	Sp012 – Sp127	Sp126 – Sp127	Sp180 – Sp129	Sp180 – Sp082	Sp129 – Sp082
F1	+	+	+	–	–	–
F2	–	–	–	+	+	–
F3	+	+	+	+	–	+
F1	+	+	–	–	–	–
F2	–	–	+	+	+	–
F3	+	+	+	+	+	+

Table 7.1: Summary of the between-speaker differences for the vowel /a/ in stressed position (row 1-3) and in unstressed position (row 4-6) in spontaneous speech. “+”:  $p < 0.05$ , “–“:  $p > 0.05$ .

/e/	Sp012 – Sp126	Sp012 – Sp127	Sp126 – Sp127	Sp180 – Sp129	Sp180 – Sp082	Sp129 – Sp082
F1	+	+	–	+	+	–
F2	–	–	–	+	+	+
F3	–	–	–	+	+	+
F1	–	–	–	–	–	–
F2	–	+	+	–	+	+
F3	–	+	+	+	+	+

Table 7.2: Summary of the between-speaker differences for the vowel /e/ in stressed position (row 1-3) and in unstressed position (row 4-6) in spontaneous speech. “+”:  $p < 0.05$ , “–“:  $p > 0.05$ .

/ɛ/	Sp012 – Sp126	Sp012 – Sp127	Sp126 – Sp127	Sp180 – Sp129	Sp180 – Sp082	Sp129 – Sp082
F1	+	–	–	+	+	+
F2	–	–	–	+	+	–
F3	+	+	+	+	+	–
F1	+	+	–	+	+	–
F2	–	–	–	+	+	–
F3	+	+	+	+	+	–

<sup>158</sup> Too few items of front rounded vowels occurred in the data for a statistical analysis.

Table 7.3: Summary of the between-speaker differences for the vowel /ɛ/ in stressed position (row 1-3) and in unstressed position (row 4-6) in spontaneous speech. “+”:  $p < 0.05$ , “-“:  $p > 0.05$ .

/i/	Sp012 – Sp126	Sp012 – Sp127	Sp126 – Sp127	Sp180 – Sp129	Sp180 – Sp082	Sp129 – Sp082
F1	+	+	-	+	+	-
F2	-	-	-	-	-	-
F3	-	+	-	-	+	+
F1	-	-	-	+	+	-
F2	-	-	-	-	+	+
F3	-	+	+	+	+	+

Table 7.4: Summary of the between-speaker differences for the vowel /i/ in stressed position (row 1-3) and in unstressed position (row 4-6) in spontaneous speech. “+”:  $p < 0.05$ , “-“:  $p > 0.05$ .

/i/	Sp012 – Sp126	Sp012 – Sp127	Sp126 – Sp127	Sp180 – Sp129	Sp180 – Sp082	Sp129 – Sp082
F1	-	-	-	+	+	-
F2	-	-	-	-	+	+
F3	-	-	-	+	+	-
F1	+	+	-	+	+	-
F2	-	-	-	-	+	+
F3	-	-	-	+	+	+

Table 7.5: Summary of the between-speaker differences for the vowel /i/ in stressed position (row 1-3) and in unstressed position (row 4-6) in spontaneous speech. “+”:  $p < 0.05$ , “-“:  $p > 0.05$ .

/u/	Sp012 – Sp126	Sp012 – Sp127	Sp126 – Sp127	Sp180 – Sp129	Sp180 – Sp082	Sp129 – Sp082
F1	-	-	-	-	-	-
F2	-	-	-	+	+	+
F3	-	-	-	+	+	+
F1	-	-	-	-	-	-
F2	-	-	-	-	-	-
F3	-	-	-	-	-	-

Table 7.6: Summary of the between-speaker differences for the vowel /u/ in stressed position (row 1-3) and in unstressed position (row 4-6) in spontaneous speech. “+”:  $p < 0.05$ , “-“:  $p > 0.05$ .

/ʊ/	Sp012 – Sp126	Sp012 – Sp127	Sp126 – Sp127	Sp180 – Sp129	Sp180 – Sp082	Sp129 – Sp082
F1	NA	NA	–	–	–	–
F2	NA	NA	–	–	–	–
F3	NA	NA	+	–	–	–
F1	NA	NA	–	+	+	–
F2	NA	NA	–	–	+	–
F3	NA	NA	–	+	–	–

Table 7.7: Summary of the between-speaker differences for the vowel /ʊ/ in stressed position (row 1-3) and in unstressed position (row 4-6) in spontaneous speech. “+”:  $p < 0.05$ , “–”:  $p > 0.05$ , NA = not enough data available.

/o/	Sp012 – Sp126	Sp012 – Sp127	Sp126 – Sp127	Sp180 – Sp129	Sp180 – Sp082	Sp129 – Sp082
F1	+	+	–	+	+	–
F2	–	–	–	–	–	–
F3	+	+	+	+	+	–
F1	–	–	–	+	+	–
F2	–	–	–	–	+	–
F3	–	+	–	+	–	–

Table 7.8: Summary of the between-speaker differences for the vowel /o/ in stressed position (row 1-3) and in unstressed position (row 4-6) in spontaneous speech. “+”:  $p < 0.05$ , “–”:  $p > 0.05$ , NA = not enough data available.

/ɔ/	Sp012 – Sp126	Sp012 – Sp127	Sp126 – Sp127	Sp180 – Sp129	Sp180 – Sp082	Sp129 – Sp082
F1	–	–	–	–	+	+
F2	–	–	–	+	+	+
F3	–	+	+	–	+	+
F1	NA	NA	–	+	+	–
F2	NA	NA	–	–	+	–
F3	NA	NA	+	–	+	–

Table 7.9: Summary of the between-speaker differences for the vowel /ɔ/ in stressed position (row 1-3) and in unstressed position (row 4-6) in spontaneous speech. “+”:  $p < 0.05$ , “–”:  $p > 0.05$ , NA = not enough data available.

The results are slightly better for the female speakers who can be discriminated with respect to the vowels /a/, /e/, /ɪ/, and /u/ in stressed position and /a/, /e/, /i/, /ɪ/ in

unstressed position. The male speakers are discriminated with respect to the vowels /a/, /ɛ/, and /o/ in stressed position, and /a/ and /ɛ/ in unstressed position. Although F3 has the highest discriminatory power, as has already be stated by others (Jessen 1997), the results of F3 are still too unreliable for discrimination. The vowel which renders a consistent result is the vowel /a/, followed by some front vowels. However, as concerns the other vowels, one never knows how they will behave with respect to speaker discrimination. Therefore, as long as one has to rely on casuistry, vowels cannot be pooled for speaker discrimination.

This holds also for vowels controlled for phonetic context. Figure 7.2 presents the density plot for all /i/ – vowels from the word “Wien” (Vienna).

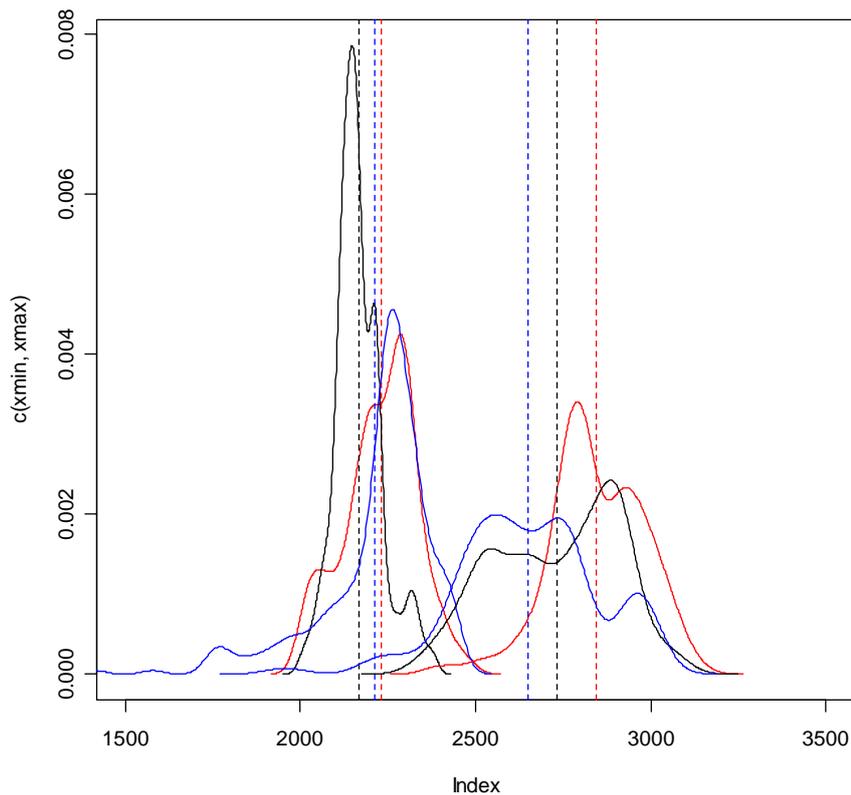


Figure 7.2: Density plot of F2 and F3 of the vowel /i/ from “Wien” (Vienna), spontaneous speech. Red: speaker sp012, black: speaker sp126, blue: speaker sp127, n = 22.

Although Figure 7.2 suggests significant differences between the three speakers by means of F3, these differences cannot withstand a statistical analysis. Neither a two-tailed t-test nor the z-deviation renders differences for F3 between sp126 and sp127.

One could argue that a visual comparison of the distributions might suffice to keep the speakers apart. However, in a cross-check comparison of one and the same speaker in different speaking tasks, the distribution of formants need not result in a satisfying overlap either. Figure 7.3 shows a density plot of all vowels /i/ from “nicht” (not) of one speaker (sp126) in the sentence reading task and in spontaneous speech:

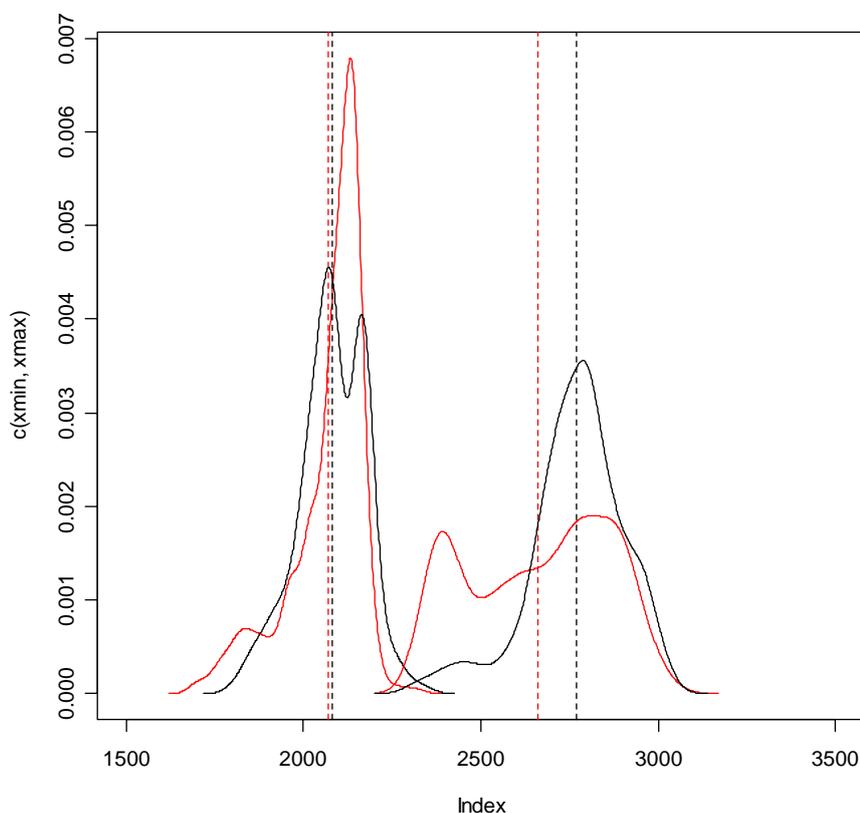


Figure 7.3: Density plot of F2 and F3 of the vowel /i/ from “nicht” (not), sentence reading task (red) and spontaneous speech (black), n = 18.

A visual inspection would render differences for F3. Calculated z-scores, however, correctly prove accordance for the /i/ of spontaneous speech with the /i/ of the sentence reading task with respect to both F2 and F3.

It might be argued that a sufficient number (~ 3-4) of several identical items (~ 20) in both the questioned and the reference material would lead to a successful comparison. Apart from the fact that in a true forensic case, a sufficient number of several identical items are hardly ever available, this hypothesis does not bear up either. Table 7.10 shows the results of the statistical analysis of a comparison of several items which were available in the spontaneous speech of all three male speakers.

Item	Sp012 – Sp126	Sp012 – Sp127	Sp126 – Sp127
geboren (born: PP)	–	–	–
aufgewachsen (grew up (PP)	+	+	+
bin (be: 1P)	+	+	–
mit (with)	–	–	–
dann (then)	–	–	–
hat (has: 3P)	+	NA	NA
hab (have: 1P)	NA	NA	–

Table 7.10: Results of the statistical analysis (F2 and F3) for vowels in identical phonetic context, spontaneous speech. “+”:  $p < 0.05$  for either F2 or F3 or both, “–”:  $p > 0.05$ , NA = not enough data available.

As can be seen from Table 7.10, even the discriminatory power of identical items is rather poor. Therefore, it is not only necessary to control the phonetic context, which actually means the word, but to additionally control the position within the sentence or utterance. It has already been argued in 6.5 that the clearer the context is defined, the sooner invariant patterns will be observed. This finding, as well as the results on prosody, imply some space for speaker-specific handling of variability. Therefore, an acoustic analysis of the same word in the same position will the soonest and the most frequently be able to differentiate speakers and to recognize identical speakers. Figure 7.4 presents the cluster analysis of the vowel /i/ from “die” (the:DEF):

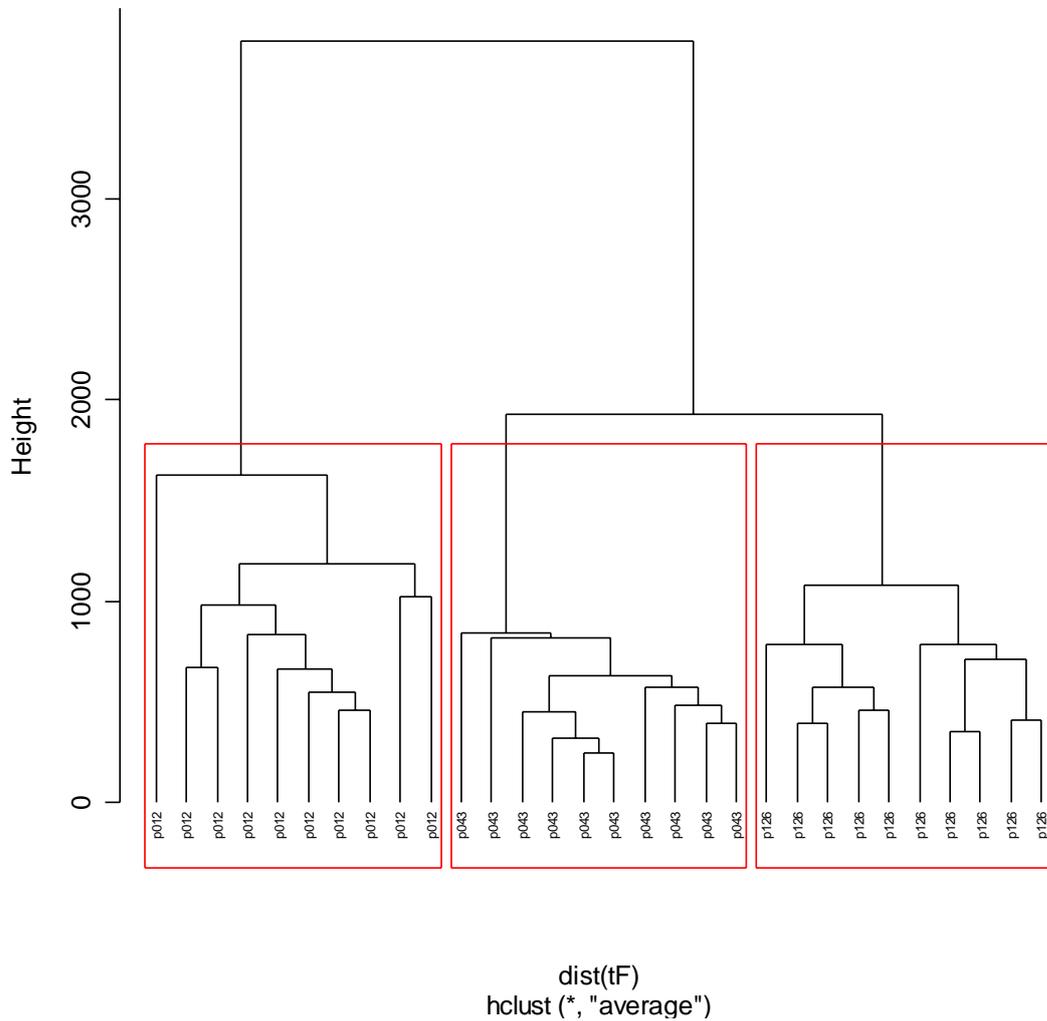


Figure 7.4: Cluster analysis of the vowel /i/ (F1, F2, F3) from “die” (the:DEF) for three male speakers, sentence repeating task.

This example is from the sentence repeating task, where speakers had to repeat one and the same sentence ten times. Therefore, the vowels are maximally controlled for context. And, as can be read from Figure 7.4, the three speakers can be differentiated. Table 7.11 presents the results of the statistical analysis of several items from the sentence repeating task.

Item	Sp012 – Sp043	Sp012 – Sp126	Sp043 – Sp126
/i/ die (the: DEF)	+	+	+
/e/ <u>g</u> eriebene (grated: PPP)	+	+	+
/i/ ger <u>i</u> ebene (grated: PPP)	+	+	–
/e/ Leber (liver)	–	+	+
/a/ Papier (paper)	+	+	+
/a/ Katzen (cats)	+	+	+

Table 7.11: Results of the statistical analysis (F2 and F3) for vowels in identical phonetic context, sentence repeating task. “+”:  $p < 0.05$  for either F2 or F3 or both, “–”:  $p > 0.05$ .

It becomes evident that the discriminatory power of F2 and F3 increase by adding the position within the sentence. Additionally, F2 and F3 discriminate unstressed vowels in every case, whereas, with the exception of the vowel /a/, F2 and F3 do not discriminate stressed vowels in two cases (sp012 and sp043 for the vowel /e/, and sp043 and sp126 for the vowel /i/).

It can be concluded, therefore, that primarily the backgrounded parts of speech give some leeway for the introduction of speaker-specific traits. The more stressed a vowel is, the more speakers head towards a corporate (acoustic) output. Moreover, the results point to a thorough planning of the speech events, down to the lowest level of the phonetic output, because only by an exact planning of what one wants to say, can such high consistency with respect to context and position be achieved.

## 8. Summary and Outlook

Approximately 11.000 vowels of six speakers of Standard Austrian German were analysed in two speaking tasks (reading sentences and spontaneous speech). Two speakers additionally read a list of logatomes. The vowels have been segmented manually, and F1, F2, F3, F0, and duration (in number of periods) have been calculated and submitted to statistical analysis. Each speaker has been discussed separately.

The acoustic analysis proved that the vowels of Standard Austrian German are located at five constriction locations: pre-palatal, palatal, velar, upper pharyngeal, and lower pharyngeal. Except for the pharyngeal vowel, each location is further distinguished by two degrees of constriction. The pre-palatal and the palatal vowels are further distinguished by lip protrusion, building a vowel system of thirteen vowels. The vowels are distinguished by the following features: [±constricted], [±open], [±round], [±front], [±lower pharyngeal], [±velar], and [±pre-palatal]. The feature [±open] was added because speaker-specific differences occurred in weak prosodic positions. Some speakers discerned the vowels by degree of openness, others by degree of constriction. Duration did not prove to be discriminatory. The feature [±tense] was abandoned in order to avoid misleading semantic implications which have no articulatory basis. Tendencies towards neutralization could be observed for the vowel pairs /i/-/ɪ/, /y/-/ʏ/, and /u/-/ʊ/. One speaker also neutralized the opposition /e/-/ɛ/ in the unstressed positions of spontaneous speech. The results on F0 did not corroborate the frequently observed correlation of tongue height and F0. The same holds for the correlation of tongue height and duration.

It is especially those tendencies towards neutralisation that should be taken note of in further investigations on the vowels of Standard Austrian German. Standard Austrian German stands in strong interaction with the Middle-Bavarian dialects, which do not

distinguish their vowels by the feature [±constricted], but rather by the feature [±long]<sup>159</sup> (Dressler/Wodak 1982, Moosmüller 1987). Therefore, a sound change in progress might be at work, changing the vowel qualities of the vowels of Standard Austrian German. This change in quality will affect the feature [±constricted] in the way that the [–constricted] vowels /i/, /y/, and /u/ become [+constricted]. By such a change, not only the vowel system of Standard Austrian German would be reduced, but, concurrently, the pre-palatal constriction location would be given up (as could already be observed for the youngest speaker, sp127). The pre-palatal location is acoustically unstable and exploited by only a few languages. Therefore, as soon as a vowel system shrinks, the palatal constriction location suffices for distinguishing the front vowels.

Phenomena usually termed as coarticulation, e.g. anticipatory lip protrusion or palatalisation, are found to be processes. In Standard Austrian German, both lip protrusion and palatalisation start at plosive release and do not affect the transconsonantal vowel, as long as the transconsonantal vowel and the plosive are separated by a word boundary.

The vowel /i/ causes a palatalisation of the preceding plosive, resulting in a plosive configuration similar to the one observed for Russian palatalised plosives. However, Russian palatalised plosives change the quality of the vowel /i/, whereas in Standard Austrian German, it is the vowel which changes the plosive. Therefore, such phenomena are to be described as processes guided by the phonology of the language.

The three places of articulation distinctive for Standard Austrian German plosives (bilabial, alveolar, and velar) could be discriminated via the transition of F2 when the preceding vowel was /a/. F2 transition discriminated the alveolar plosive preceding the vowels /u/ and /o/ from the velar and bilabial plosives. In velar and bilabial contexts, F2 is lower at vowel onset. Preceding the vowel /i/, the bilabial context was discriminated by all speakers, whilst the differentiation of alveolar and velar context was less secure.

---

<sup>159</sup> Whether a feature [±long] is to be assumed for the Viennese dialect, has still to be examined.

Speaking is understood as a social interaction. Therefore, the interactional situation determines speech behaviour to a large extent. The variability of vowels depends, extra-linguistically, on the interactional situation and, intra-linguistically, on the prosodic strength of a given vowel. It is argued that in more informal interactional situations, speakers do not adopt a relaxed speaking mode, allegedly easing articulation, but apply the processes adequate for the given interactional situation. This argumentation is proved by the examination of articulatory avoidance, where it is the plan of the speaker to depart as little as possible from the neutral vocal tract configuration. This purpose results in a completely unsystematic variation. Within a given interactional situation, variation is, however, highly systematic. Consequently, concepts like e.g. the concept of undershoot or the concept of low cost, are abandoned.

The examination of “fronting” of back vowels further corroborates this argumentation. In Standard Austrian German, F2 displacements can be observed in the most formal speaking task – the reading of logatomes – for the vowels /ʊ/ and /ɔ/. In the task of reading sentences, displacements of F2 can only be observed for the vowel /ɔ/. In spontaneous speech, the differences no longer exist. These results strongly suggest that in Standard Austrian German, F2 displacement is neither a matter of undershoot nor a coarticulatory phonetic detail, but a process maximising contrast in most formal speech situations. Therefore, in any phonetic investigation, the phonology of the analysed language has to be incorporated.

Prosodic strength is expressed by both quantity and quality. Unstressed vowels are significantly shorter than stressed vowels, in both speaking tasks. However, no correlation was observed between duration and either F1, F2, or F3. Therefore, duration plays no relevant role in the qualitative change of vowels. The change in prosodic strength affects predominantly F2, except for the vowel /a/, which shows a significantly higher F1 in unstressed positions. It has to be emphasized that a change in prosodic strength only changes the quality of the vowel, whilst the vowel category is unaffected.

Primary stressed vowels are distinguished from unstressed vowels by all speakers. However, the assignment of secondary stress turned out to be speaker-specific. Secondary stress might either be realised in the same way as the primary stress, or in the same way as the unstressed position, or as a discrete secondary stress. In the same way, sentence stress is modeled differently by the speakers, in this way assigning, even on the segmental level, a specific rhythm to utterances.

From these results, it can be concluded that the final shape of primary stressed positions and unstressed positions is regulated, and provides the speaker with a frame within which he or she can move for the modelling or not-modelling of further levels of stresses. The specific frame is defined by the interactional situation. Therefore, many speaker-specific differences can be found in those positions which allow the speaker some degrees of freedom in modelling his or her speech behaviour.

## 9. References

- Abbs, James H. 1986. "Invariance and Variability in Speech Production: A Distinction Between Linguistic Intent and Its Neuromotor Implementation." In: Joseph S. Perkell & Dennis H. Klatt (ed.), *Invariance and Variability in Speech Processes*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 202-219.
- Abry, Christian, Louis-Jean Boë & Jean-Luc Schwartz. 1989. "Plateaus, catastrophes and the structuring of vowel systems." *Journal of Phonetics* 17, 47-54.
- Adank, Patty, Roel Smits & Roeland van Hout. 2003. "Modeling perceived vowel height, advancement, and rounding." *Proceedings of the 15<sup>th</sup> International Congress on Phonetic Sciences*, Barcelona, August 3-9, 2003, 647-650.
- Al-Bamerni, A. & A. Bladon. 1982. "One-stage and two-stage patterns of velar coarticulation." *Journal of the Acoustical Society of America* 72, 104.
- Alfonso, Peter J. & Thomas Baer. 1982. "Dynamics of vowel articulation." *Language and Speech* 25, 151-173.
- Andruski, Jean E. & Terrance M. Nearey. 1992. "On the Sufficiency of Compound Target Specification of Isolated Vowels and Vowels in /bVb/ Syllables." *Journal of the Acoustical Society of America* 91 (1), 390-410.
- Antoniadis, Zissis & Hans Werner Strube. 1984. "Untersuchungen zur spezifischen Dauer deutscher Vokale." *Phonetica* 41, 72-87.
- Apostol, Lian, Pascal Perrier, Monica Raybaudi & Christoph Segebarth. 1999. "3D Geometry of the Vocal Tract and Inter-Speaker Variability." *Proceedings of the 14<sup>th</sup> International Congress on Phonetic Sciences*, San Francisco, August 1-7, 1999, 443-446.
- Atal, Bishnu S., J. J. Chang & M. V. Matthews & J. W. Tukey. 1978. "Inversion of Artiucloatry-to-acoustic Transformation in the Vocal Tract by a Computer-sorting Technique." *Journal of the Acoustical Society of America* 63, 1535-1555.
- Badin, Pierre, Pascal Perrier, Louis-Jean Boë & Christian Abry. 1990. "Vocalic nomograms: Acoustic and articulatory considerations upon formant convergences." *Journal of the Acoustical Society of America* 87 (3), 1290-1300.
- Bannert, Robert. 1977. "Quantität im Mittelbairischen: Komplementäre Länge von Vokal und Konsonant." In: Wolfgang U. Dressler, Oskar E. Pfeiffer & Thomas Herok (eds.), *Phonologica 1976*. Innsbruck: 261-270.
- Barbour, Stephen & Stevenson, Patrick. 1990. *Variation in German. A critical approach to German sociolinguistics*. Cambridge: Cambridge University Press.
- Barry, William J. & Bistra Andreeva. 2001. "Cross-language similarities and differences in spontaneous speech patterns." *Journal of the International Phonetic Association* 31 (1), 51-66.
- Bates, Sally A. R. 1995. "Acoustic characterization of schwa: a comparative study." *Proceedings of the 13<sup>th</sup> International Congress on Phonetic Sciences*. Stockholm. Vol 3, 230-233.
- Båvegård, Mats & Gunnar Fant. 1996. "Parameterized VT Area Function Inversion." 4<sup>th</sup> International Conference on Spoken Language Processing ICSLP '96, Philadelphia, PA, October 3-6, 961-964.

- Becker, Thomas. 1998. *Das Vokalsystem der deutschen Standardsprache*. Frankfurt/Main: Peter Lang.
- Behne, Dawn, Takayuki, Arai, Peter Czigler & Kirk Sullivan. 1999. "Vowel Duration and Spectra as Perceptual Cues to Vowel Quantity: A Comparison of Japanese and Swedish." *Proceedings of the 14<sup>th</sup> International Congress on Phonetic Sciences*, San Francisco, August 1-7, 1999, 857-860.
- Beckman, Mary E., Tzyy-Ping Jung, Sook-hyang Lee, Kenneth de Jong, Ashok K. Krishnamurthy, Stanley C. Ahalt, K. Bretonnel Cohen & Michael J. Collins. 1995. "Variability in the production of quantal vowels revisited." *Journal of the Acoustical Society of America* 97 (1), 471-490.
- Benguerel, A.P., H. Hirose, M. Sawashima & T. Ushijima. 1977. "Velar coarticulation in French: a fiberoptic study." *Journal of Phonetics* 5, 149-158.
- Bennett, David C. 1968. "Spectral form and duration as cues in the recognition of English and German vowels." *Language and Speech* 11, 65-85.
- Blumstein, Shiela E. 1986. "On Acoustic Invariance in Speech." In: Joseph S. Perkell & Dennis H. Klatt (ed.), *Invariance and Variability in Speech Processes*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 178-193.
- Blumstein, Shiela E. & Kenneth N. Stevens. 1979. "Acoustic Invariance in Speech Production: Evidence From Measurements of the Spectral Characteristics of Stop Consonants." *Journal of the Acoustical Society of America* 66 (4), 1001-1017.
- Blumstein, Shiela E. & Kenneth N. Stevens. 1980. "Perceptual Invariance and Onset Spectra for Stop Consonants in Different Vowel Environments." *Journal of the Acoustical Society of America* 67 (2), 648-662.
- Boë, Louis-Jean & Pascal Perrier. 1990. "Comments on „Distinctive Regions and Modes: A new Theory of Speech Production“ by M. Mrayati, René Carré and B. Guérin [Speech Communication 7(3), 257-286 (1988)]." *Speech Communication* 9, 217-230.
- Boë, Louis-Jean, Pascal Perrier & Gérard Bailly. 1992. "The geometric vocal tract variables controlled for vowel production: proposals for constraining acoustic-to-articulatory inversion." *Journal of Phonetics* 20 (1), 27-38.
- Bohn, Ocke-Schwen. 2004. "How to organize a fairly large vowel inventory: the vowels of Fering (North Frisian)." *Journal of the International Phonetic Association* 34 (2), 161-173.
- Bohn, Ocke-Schwen, James E. Flege, Paul A. Dagenais & Samuel G. Fletcher. 1992. "Differenzierung und Variabilität der Zungenpositionen bei der Artikulation deutscher Vokale." In: Wolfgang Hess & Walter F. Sendlmeier (eds.), *Beiträge zur angewandten und experimentellen Phonetik*. Stuttgart: Steiner, 1-26. (=Zeitschrift für Dialektologie und Linguistik, Beiheft 72).
- Boyce, Suzanne E. 1990. "Coarticulatory organization for lip rounding in Turkish and English." *Journal of the Acoustical Society of America* 88 (6), 2584-2595.
- Browman, Catherine P. & Louis Goldstein. 1992. "'Targetless' schwa: An articulatory analysis." In: John Kingston & Mary E. Beckman (eds.), *Papers in Laboratory Phonology II*. Cambridge: Cambridge University Press, 26-56.
- Browman, Catherine P. 1995. "Assimilation as gestural overlap: comments on Holst and Nolan." In: Bruce Connell & Amalia Arvaniti (eds.), *Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV*. Cambridge: Cambridge University Press, 334-342.
- Brunner, Jana, Susanne Fuchs & Pascal Perrier. 2005. "The influence of the palate shape on articulatory token-to-token variability." *ZAS Papers in Linguistics* 42, 43-67.
- Camaj, Martin. 1969. *Lehrbuch der albanischen Sprache*. Wiesbaden: Harrassowitz.

- Carré, René. 1996. "Prediction of Vowel Systems using a Deductive Approach." *Proceedings of the International Congress on Speech and Language Processing, Philadelphia*, 434-437.
- Carré, René. 2004. "From an acoustic tube to speech production." *Speech Communication* 42, 227-240.
- Carré, René & M. Mrayati. 1991. "Vowel-vowel trajectories and region modeling." *Journal of Phonetics* 19, 433-443.
- Carré, René & Maria Mody. 1997. "Prediction of Vowel and Consonant Place of Articulation." *Proceedings of the third Meeting of the ACL Interest Group in Computational Phonology, SIGPHON 97*, 26-32.
- Carré, René & Pierre L. Divenyi. 2000. Modeling and Perception of 'Gesture Reduction'. *Phonetica* 57, 152-169.
- Chafcouloff, Michel & Alain Marchal. 1999. "Velopharyngeal coarticulation." In: William J. Hardcastle & Nigel Hewlett (eds.), *Coarticulation. Theory, Data and Techniques*. Cambridge: Cambridge University Press, 69-79.
- Cho, Taehong. 2004. "Prosodically-Conditioned Strengthening and Vowel-to-Vowel Coarticulation in English." *Journal of Phonetics* 32 (2), 141-176.
- Cho, Taehong & Patricia A. Keating. (2001). "Articulatory and acoustic studies on domain-initial strengthening in Korean." *Journal of Phonetics* 29, 155-190.
- Cho, Taehong & James McQueen. 2005. "Prosodic influences on consonant production in Dutch: Effects of prosodic boundaries, phrasal accent and lexical stress." *Journal of Phonetics* 33 (2), 121-157.
- Chomsky, Noam & Morris Halle. 1968. *The Sound Pattern of English*. New York: Harper & Row.
- Cläßen, Kathrin, Grzegorz Dogil, Michael Jessen, Krzysztof Marasek & Wolfgang Wokurek. 1998. "Stimmqualität und Wortbetonung im Deutschen." *Linguistische Berichte* 174, 202-245.
- Clumeck, Harold. 1976. "Patterns of soft palate movements in six languages" *Journal of Phonetics* 4, 337-351.
- Cohn, Abigail C. 1993. "Nasalisation in English: phonology or phonetics." *Phonology* 10, 43 – 81.
- Coulmas, Florian. 2000. "Die Frage des Standards." *Grazer Linguistische Monographien* 11, 21-34.
- Cox, Felicity. 1999. "Vowel Change in Australian English." *Phonetica* 56, 1-27.
- Crystal, Thomas H. & Arthur S. House. 1988a. "Segmental durations in connected-speech signals: Syllabic stress." *Journal of the Acoustical Society of America* 83 (4), 1574-1585.
- Cummins, Fred. 2002. "Speech Rhythm and Rhythmic Taxonomy." In: Bel, Bernard & Isabelle Marlien (eds.), *Proceedings of Speech Prosody 2002*, 121-126.
- Dang, Jianwu, Jianguo Wei, Takeharu Suzuki, Kiyoshi Honda, Pascal Perrier & Masaaki Honda. 2004. "Investigation and Modeling of Coarticulation in Speech Production." 4<sup>th</sup> International Symposium on Chinese Spoken Language Processing ISCSLP 2004, 25-28.
- Daniiloff, Raymond G. & Kenneth L. Moll. 1968. "Coarticulation of lip rounding" *Journal of Speech and Hearing Research* 11, 707-721.
- Davidson, Lisa & Maureen Stone. 2003. "Epenthesis Versus Gestural Mistiming in consonant Cluster Production: an Ultrasonic Study." *WCCFL* 22, 165-178.
- Dembovsky, James, Mary J. Lindstrom & John R. Westbury. 1998. "Articulator Point Variability in the Production of Stop Consonants." In: Michael P. Cannito, Kathryn M. Yorkston & David R. Beuckelman (eds.), *Neuromotor speech disorders: Nature*

- assessment and management. Baltimore: Paul H. Brookes Publishing Co., 27-46.
- Diehl, Randy L., Björn Lindblom & Carl P. Creeger. 2003. "Increasing Realism of Auditory Representations Yields further Insight into Vowel Phonetics." *Proceedings of the 15<sup>th</sup> International Congress on Phonetic Sciences, Barcelona, August 3-9, 2003*, 1381-1384.
- Diehl, Randy L. & Björn Lindblom. 2004. "Explaining the Structure of Feature and Phoneme Inventories: The Role of Auditory Distinctiveness." In: Steven Greenberg, Arthur N. Popper, William A. Ainsworth & Richard R. Fay (eds.), *Speech Processing in the Auditory System*. New York: Springer, 101-162.
- Dilley, Laura, Stefanie Shattuck-Hufnagel & Mari Ostendorf. (1996). "Glottalization of word-initial vowels as a function of prosodic structure." *Journal of Phonetics* 24, 423-444.
- Disner, Sandra F. 1986. "On Describing Vowel Quality." In: John J. Ohala & Jeri J. Jaeger (eds.), *Experimental Phonology*. Orlando, FL: Academic Press, 69-79.
- Dittmar, Norbert. 1997. *Grundlagen der Soziolinguistik*. Tübingen: Niemeyer.
- Dogil Grzegorz & Briony Williams. 1999. "The phonetic manifestation of word stress." In: Harry van der Hulst (ed.), *Word Prosodic Systems in the Languages of Europe*. Berlin, New York: Mouton de Gruyter, 273-334.
- Doleschal, Ursula. 1988. "Zum deutschen Kompositionsakzent. Tema con variazioni." *Wiener Linguistische Gazette* 40-41, 3-27.
- Donegan, Patricia J. 1995. "The innateness of phonemic perception." In: Vida Samiian (ed.), *Proceedings of the 24<sup>th</sup> Western Conference on Linguistics (WECOL 7)*. Fresno, CA: Western Conference on Linguistics, 59-69.
- Donegan, Patricia J. 2002. "Phonological Processes and Phonetic Rules." In: Katarzyna Dziubalska-Kolaczyk & Jarek Weckwerth (eds.), *Future Challenges for Natural Linguistics*. München: Lincom Europa, 57-81.
- Donegan, Patricia J. & David Stampe. 1979. "The Study of Natural Phonology." In: David A. Dinnsen (ed.), *Current Approaches to Phonological Theory*. Bloomington & London: Indiana University Press, 126-173.
- Deutsch, Werner A., Sylvia Moosmüller & Ralf Vollmann. 1994. "Ausdruck von Stimme und Sprechweise nach Felix Trojan." In: Karl Fellbaum (ed.), *Elektronische Sprachsignalverarbeitung. Fünfte gemeinsame Konferenz der Technischen Universitäten Berlin und Dresden und der Humboldt-Universität zu Berlin*. Berlin, 5.-7. Oktober 1994. Berlin: TU Berlin, Institut für Fernmeldetechnik (= Studientexte zur Sprachkommunikation 11), 151-157.
- Dressler, Wolfgang U. 1979. "Reflections on phonological typology." *Acta Linguistica Academiae Scientiarum Hungaricae*, Tomus 29 (3-4), 259-273.
- Dressler, Wolfgang U. 1980. "A Semiotic Model of Diachronic Process Phonology." *Wiener Linguistische Gazette* 22-23, 30-94.
- Dressler, Wolfgang U. 1984. "Explaining natural phonology." *Phonology Yearbook* 1, 29-50.
- Dressler, Wolfgang U. 1985. *Morphonology*. Ann Arbor: Karoma.
- Dressler, Wolfgang U. 1996. "Principles of naturalness in phonology and across components." In: Bernhard Hurch & Richard A. Rhodes (eds.), *Natural Phonology: The State of the Art*. Berlin, New York: Mouton de Gruyter, 41-51.
- Dressler, Wolfgang U. 2002. "Naturalness and functionalism." In: Katarzyna Dziubalska-Kolaczyk & Jarek Weckwerth (eds.), *Future Challenges for Natural Linguistics*. München: Lincom Europa, 83-101.
- Dressler, Wolfgang U. & Ruth Wodak. 1982. "Sociophonological methods in the study of sociolinguistic variation in Viennese German." *Language in Society* 11, 339-370.

- Dressler, Wolfgang U. & Sylvia Moosmüller. 1991. "Phonetics and Phonology: A Sociopsycholinguistic Framework." *Phonetica* 48, 135-148.
- Ejstrup, Michael & Gert Foget Hansen. 2003. "Danish Vowels in Spontaneous Speech in Three Modern Regional Variants." *Proceedings of the 15<sup>th</sup> International Congress on Phonetic Sciences, Barcelona, August 3-9, 2003*, 2119-2122.
- Engstrand, Olle. 1988. "Articulatory Correlates of Stress and Speaking Rate in Swedish VCV Utterances." *Journal of the Acoustical Society of America* 83 (5), 1863-1875.
- Engstrand, Olle & Diana Krull. 2001. „Segment and syllable reduction: preliminary observations.“ *Lund Working Papers* 49, 26-29.
- Erickson, Donna. 2002. „Articulation of extreme formant patterns for emphasized vowels“.  
*Phonetica* 59, 134-149.
- Ericsson, Christine, Björn Lindblom & Johan Stark. 1999. "Articulatory coordination in coronal stops: Implications for theories of coarticulation." *Proceedings of the 14<sup>th</sup> International Congress on Phonetic Sciences, San Francisco, August 1-7, 1999*, 1885-1888.
- Fant, Gunnar. 1964. "Phonetics." In: David E. Brewer (ed.), *Research Potentials in Voice Physiology*. New York: State University of New York, 197-239.
- Fant, Gunnar. 1965. „Formants and Cavities.“ *Proceedings of the 5<sup>th</sup> International Congress on Phonetic Sciences, Münster 1964*, 120-141.
- Fant, Gunnar. 1970<sup>2</sup>. *Acoustic Theory of Speech Production*. The Hague: Mouton.
- Fant, Gunnar. 1986. "Features: Fiction and Facts." In: Joseph S. Perkell & Dennis H. Klatt (eds.), *Invariance and Variability in Speech Processes*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 480-492.
- Fant, Gunnar. 2001. "Swedish vowels and a new three-parameter model." *TMH-QPSR* 1/2001, 43-49.
- Fant, Gunnar. 2004. "The Relations between Area Functions and the Acoustical Signal" In: Gunnar Fant (ed.), *Speech Acoustics and Phonetics. Selected Writings*. Dordrecht, Boston, London: Kluwer Academic Publishers, 29-57. (= 1980, *Phonetica* 37, 55-86).
- Fant, Gunnar & Stefan Pauli. 1974. "Spatial Characteristics of Vocal Tract Resonance Modes." In: Gunnar Fant (ed.), *Speech Communication 2. Speech Production and Synthesis by Rules*. Stockholm: Almqvist & Wiksell International, 121-133.
- Fant, Gunnar, Lennart Nord & Anita Kruckenberg. 1986. "Individual variations in text reading. A databank pilot study." *STL-QPSR* 1986, 1-17.
- Fant, Gunnar & Mats Båvegård. 1997. „Parametric model of VT area functions. vowels and consonants“ *THM-QPSR* 1, 1-21.
- Fant, Gunnar, Anita Kruckenberg & Johan Liljencrants. 2000a. "Acoustic-phonetic Analysis of Prominence in Swedish." In: Antonis Botinis (ed.), *Intonation. Analysis, Modelling and Technology*. Dordrecht, Boston, London: Kluwer Academic Publishers, 55-86.
- Fant, Gunnar, Anita Kruckenberg & Johan Liljencrants. 2000b. "The Source-Filter Frame of Prominence". *Phonetica* 57, 113-127.
- Fant, Gunnar, Anita Kruckenberg, Johan Liljencrants & Antonis Botinis. 2002. "Individual variations in prominence correlates. Some observations from „lab-speech“." *THM-QPSR* 44, 177-180.
- Fant, Gunnar & Anita Kruckenberg. 2004. "An Integrated View of Swedish Prosody. Voice Production, Perception and Synthesis." In: Gunnar Fant (ed.), *Speech Acoustics and Phonetics. Selected Writings*. Dordrecht, Boston, London: Kluwer Academic Publishers, 249-300.
- Farnetani, Edda. 1990. "V-C-V Lingual Coarticulation and its Spatiotemporal Domain." In:

- William J. Hardcastle & Alain Marchal (eds.), *Speech Production and Speech Modelling*. Dordrecht, Boston, London: Kluwer Academic Publishers, 93-130.
- Farnetani, Edda. 1991. "Coarticulation and reduction in consonants: comparing isolated words and continuous speech." *PERILUS XIV*, 11-15.
- Farnetani, Edda. 1997. "Coarticulation and Connected Speech Processes." In: William J. Hardcastle & John Laver (eds.), *The Handbook of Phonetic Sciences*. Oxford and Cambridge: Blackwell, 371-404.
- Farnetani, Edda. 1999. "Labial coarticulation." In: William J. Hardcastle & Nigel Hewlett (eds.), *Coarticulation. Theory, Data and Techniques*. Cambridge: Cambridge University Press, 144-163.
- Farnetani, Edda & Daniel Recasens. 1993. "Anticipatory Consonant-to-Vowel Coarticulation in Italian." *Language and Speech* 36, 279-302.
- Farnetani, Edda & Daniel Recasens. 1999. "Coarticulation models in recent speech production theories." In: William J. Hardcastle & Nigel Hewlett (eds.), *Coarticulation. Theory, Data and Techniques*. Cambridge: Cambridge University Press, 31-65.
- Féry, Caroline & Laura Herbst. 2004. "German Sentence Accent Revisited." In: Shinichiro Ishihara, Michaela Schmitz & Anne Schwarz (ed.), *Interdisciplinary studies on information structure*, Vol. 1. Potsdam: Univ.-Verlag, 43-75.
- Fisher-Jørgensen, Eli. 1990. Intrinsic F0 in Tense and Lax Vowels with Special Reference to German. *Phonetica* 47 (3-4), 99-140.
- Flege, James E. 1989. "Differences in inventory size affect the location but not the precision of tongue positioning in vowel production." *Language and Speech* 32 (2), 123-147.
- Flemming, Edward. 2001. „Scalar and categorical phenomena in a unified model of phonetics and phonology.“ *Phonology* 18 (1), 7 – 44.
- Fougeron, Cécile. 2001. "Articulatory properties of initial segments in several prosodic constituents in French." *Journal of Phonetics* 29, 109-135.
- Fougeron, Cécile & Patricia A. Keating. 1997. "Articulatory strengthening at edges of prosodic domains." *Journal of the Acoustical Society of America* 101, 3728-3740.
- Fourakis, M. 1991. "Tempo, Stress, and Vowel Reduction in American English." *Journal of the Acoustical Society of America* 90, 1816-1827.
- Fowler, Carol A. 1990. "Some regularities in speech are not consequences of formal rules: comments on Keating's paper." In: John Kingston & Mary E. Beckman (eds.), *Papers in Laboratory Phonology I. Between Grammar and Physics of Speech*. Cambridge: Cambridge University Press, 476-489.
- Fowler, Carol A. & Lawrence Brancazio. 2000. "Coarticulation Resistance of American English Consonants and its Effects on Transconsonantal Vowel-to-Vowel Coarticulation." *Language and Speech* 43 (1), 1-41.
- Fuchs, Susanne, Phil Hoole, Jana Brunner & Miki Inoue. 2004. "The Trough Effect – an aerodynamic phenomenon?" *From Sound to Sense Meeting at MIT*, June 11 – June 13, 2004, C25–C30.
- Fuchs, Susanne & Pascal Perrier. 2005. "On the complex nature of speech kinematics." *ZAS Papers in Linguistics* 42, 137-165.
- Fuchs, Susanne, Pascal Perrier, Christian Geng & Christine Mooshammer. 2006. "What Role Does the Palate Play in Speech Motor Control? Insights from Tongue Kinematics for German Alveolar Obstruents." In: Jonathan Harrington & Marija Tabain (eds.), *Speech Production. Models, Phonetic Processes, and Techniques*. New York and Hove: Psychology Press, 149-164.

- Fulop, Sean A., Ethelbert E. Kari & Peter Ladefoged. 1998. „An Acoustic Study of the Tongue Root Contrast in Degema Vowels.“ *Phonetica* 55, 80 – 98.
- Galves, Antonio, Jesus Garcia, Denise Duarte & Charlotte Galves. 2002. "Sonority as a Basis for Rhythmic Class Discrimination." *Proceedings of Speech Prosody, Aix en Provence, France, April 11-13, 2002*, 323-326.
- Gay, Thomas. 1978. "Effect of speaking rate on vowel formant movements." *Journal of the Acoustical Society of America* 63 (1), 223-230.
- Gay, Thomas, Louis-Jean Boë, Pascal Perrier, Gang Feng & E. Swayne. 1991. "The acoustic sensitivity of vocal tract constrictions: Analysis by simulation." *Journal of Phonetics* 19 (3/4), 445-452.
- Gay, Thomas, Louis-Jean Boë & Pascal Perrier. 1992. "Acoustic and perceptual effects of changes in vocal tract constrictions for vowels." *Journal of the Acoustical Society of America* 92 (3), 1301-1309.
- Gick, Brian. 2002. "An X-Ray Investigation of Pharyngeal Constriction in American English Schwa." *Phonetica* 59, 38-48.
- Gick, Brian, A. Mion Kang & Douglas H. Whalen. 2000. "MRI and X-Ray Evidence for Commonality in the Dorsal Articulations of English Vowels and Liquids." *Proceedings of the 5<sup>th</sup> Seminar on Speech Production, Munich, Ludwig Maximilians-Universität*, 69-72.
- Gick, Brian & Ian Wilson. 2001. "Pre-liquid excrescent schwa: What happens when vocalic targets conflict." *Proceedings of the 7<sup>th</sup> European Conference on Speech Communication and Technology. Eurospeech 2001. Scandinavia*, 273-276.
- Gick, Brian & Ian Wilson. 2005. "Excrescent schwa and vowel laxing: Cross-linguistic responses to conflicting articulatory targets." In: Louis Goldstein, Douglas H. Whalen & Catherine T. Best (eds.), *Papers in Laboratory Phonology VIII: Varieties of Phonological Competence*. Berlin, New York: Mouton de Gruyter, 635-659.
- Goldbeck, Thomas. P. & Walter F. Sendlmeier. 1988. "Wechselbeziehung zwischen Satzmodalität und Akzentuierung in satzfinaler Position bei der Realisierung von Intonationskonturen." In: Hans Altmann (ed.), *Intonationsforschungen*. Tübingen: Niemeyer, 305-321.
- Gordon, Matthew & Ian Maddieson. 1999. "The Phonetics of Ndumbea". *Oceanic Linguistics* 38, 66-90.
- Grabe, Esther & Ee Ling Low. 2002. "Durational Variability in Speech and the Rhythm Class Hypothesis." In: Carlos Gussenhoven & Natasha Warner (eds.), *Laboratory Phonology 7: Phonology & Phonetics*. Berlin: Mouton de Gruyter, 515-546.
- Granser, Theodor & Sylvia Moosmüller. 2002. "Phonemic Schwa – a challenge for Natural Phonology? The case of Albanian." In: Katarzyna Dziubalska-Kołaczyk & Jarek Weckwerth (eds.), *Future Challenges for Natural Linguistics*. München: Lincom Europa, 165-178.
- Greenberg, Steven, Hannah Carvey, Leah Hitchcock & Shawn Chang. 2003. "Temporal properties of spontaneous speech – a syllable-centric perspective." *Journal of Phonetics* 31, 465-485.
- Guenther, Frank H. 1995. "Speech Sound Acquisition, Coarticulation, and Rate Effects in a Neural Network Model of Speech Production." *Psychological Review* 102, 594-621.
- Guenther, Frank H. 2003. "Neural control of speech movements." In: Niels O. Schiller & Antje S. Meyer (eds.), *Phonetics and Phonology in Language Comprehension and Production. Differences and Similarities*. Berlin, New York: Mouton de Gruyter, 209-239.
- Guenther, Frank H., Satrajit S. Gosh, Alfonso Nieto-Castanon & Jason A. Tourville. 2006. "A

- Neural Model of Speech Production." In: Jonathan Harrington & Marija Tabain (eds.), *Speech Production. Models, Phonetic Processes, and Techniques*. New York and Hove: Psychology Press, 27-39.
- Guion, Susan G., Mark W. Post & Doris L. Payne. 2004. "Phonetic Correlates of Tongue Root Vowel Contrasts in Maa." *Journal of Phonetics* 32 (4), 517-542.
- Haar, Carel ter. 2001. "Die niederländische Hochsprache." In: Konrad Ehlich, Jakob Ossner & Harro Stammerjohann (eds.), *Hochsprachen in Europa. Entstehung, Geltung, Zukunft*. Freiburg im Breisgau: Fillibach. 123-144.
- Halle, Morris & Kenneth N. Stevens. 1969. "On the feature Advanced Tongue Root". *Quarterly Progress Report 94*, Research Laboratory of Electronics, Massachusetts Institute of Technology, 209-215.
- Handbook of the International Phonetic Association*. 1999. Cambridge: Cambridge University Press.
- Harrington, Jonathan & Steve Cassidy. 1999. *Techniques in Speech Acoustics*. Dordrecht, Boston, London: Kluwer.
- Harrington, Jonathan, Janet Fletcher & Mary Beckman. 2000. "Manner and place conflicts in the articulation of accent in Australian English." In: Micheael B. Broe & Janet Pierrehumbert (eds.), *Language acquisition and the lexicon: Papers in Laboratory Phonology V*. Cambridge: Cambridge University Press, 40-51.
- Harshman, Richard, Peter Ladefoged & Louis Goldstein. 1977. "Factor Analysis of tongue shapes." *Journal of the Acoustical Society of America* 62, 693-707.
- Häusler, Frank. 1976<sup>2</sup>. *Das Problem der Phonetik und Phonologie bei Baudouin de Courtenay und in seiner Nachfolge*. Halle: Max Niemeyer Verlag.
- Heike Georg. 1961. "Das phonologische System des Deutschen als binäres Distinktionssystem". *Phonetica* 6, 162-176.
- Heldner, Mattias. 2003. "On the reliability of overall intensity and spectral emphasis as acoustic correlates of focal accents in Swedish." *Journal of Phonetics* 31, 39-62.
- Heldner, Mattias & Eva Strangert. 2001. "Temporal effects of focus in Swedish." *Journal of Phonetics* 29 (3), 329-361.
- Hellwag, C. F. 1781. *Dissertatio inauguralis physicomédica de formatione loquelae*. Tübingen (cited in Stratka 1978)
- Hertrich, Ingo & Hermann Ackermann. 1997. "Articulatory control of phonological vowel length contrasts: Kinematic analysis of labial gestures." *Journal of the Acoustical Society of America* 102 (1), 523-536.
- Hess, Susan. 1992. "Assimilatory effects in a vowel harmony system: an acoustic analysis of advanced tongue root in Akan." *Journal of Phonetics* 20, 475-492.
- Hillenbrand, James M. & Terrance M. Nearey. 1999. "Identification of resynthesized /hVd/ utterances: Effects of formant contour." *Journal of the Acoustical Society of America* 105 (6), 3509-3523.
- Hillenbrand, James M., Michael J. Clark & Terrance M. Nearey. 2001. "Effects of consonant environment on vowel formant patterns." *Journal of the Acoustical Society of America* 109 (2), 748-763.
- Hirata, Yukari. 2004. "Relational Acoustic Invariance in Japanese Vowel Length Distinction". *Proceedings of From Sound to Sense: Fifty +Years of Discoveries in Speech Communication*, June 11-13, 2004 at MIT, C-145-C-150.
- Hirata, Yukari & Kimiko Tsukada. 2004. "The Effects of Speaking Rates and Vowel Length on Formant Movements in Japanese." *Proceedings of the 2003 Texas Linguistics Society*

- Conference, ed. Augustine Agwuele, Willis Warren, & Sang-Hoon Park, Somerville, MA: Cascadilla Proceedings Project, 73.-85.
- Honda, Kiyoshi & Osamu Fujimura. 1991. "Intrinsic Vowel F0 and Phrase-Final Lowering: Phonological versus Biological Explanations." In: Jan Gauffin & Britta Hammarberg (eds.), *Vocal Fold Physiology: Acoustic, Perceptual, and Physiological Aspects of Voice Mechanisms*. San Diego: Singular Press, 149-157.
- Holst, Tara & Francis Nolan. 1995. "The influence of syntactic structure on [s] to [ʃ] assimilation." In: Bruce Connell & Amalia Arvaniti (eds.), *Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV*. Cambridge: Cambridge University Press, 315-333.
- Hoole, Philip. 1987. "Bite-block speech in the absence of oral sensibility." *Proceedings of the 11<sup>th</sup> International Congress on Phonetic Sciences, Tallinn, August 1-7, 1987, Vol. 4*, 16-19.
- Hoole, Philip. 1999. "Articulatory-Acoustic Relations in German Vowels." *Proceedings of the 14<sup>th</sup> International Congress on Phonetic Sciences, San Francisco, August 1-7, 1999*, 2153-2156.
- Hoole, Philip, Stefan Gfroerer & Hans G. Tillmann. 1990. "Electromagnetic articulography as a tool in the study of lingual coarticulation." *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München* 28, 107-122.
- Hoole, Philip, Christine Mooshammer & Hans G. Tillmann. 1994. "Kinematic Analysis of Vowel Production in German." *Proceedings of the ICSLP '94, Yokohama, Vol 1*, 53-56.
- Hoole, Philip & Barbara Kühnert. 1996. "Tongue-jaw coordination in German vowel production." *Proceedings of the 1<sup>st</sup> ESCA tutorial and research workshop on Speech Production Modelling/4<sup>th</sup> Speech Production Seminar*. Atrants, 97-100.
- Hoole, Philip, Kevin Munhall & Christine Mooshammer. 1998. "Do air-stream mechanisms influence tongue movement paths?" *Phonetica* 55 (3), 131 – 146.
- Hoole, Philip & Christine Mooshammer. 2002. "Articulatory analysis of the German vowel system." In: Peter Auer, Peter Gilles & Helmut Spiekermann (eds.), *Silbenschnitt und Tonakzente*. Tübingen: Niemeyer, 129-152.
- Howie, Stephen M. 2001. "Formant Transitions of Russian Palatalized and Nonpalatalized Syllables." *IULC Working Papers Online* 01-01.
- Hu, Fang. 2003. "An Acoustic and Articulatory Analysis of Vowels in Ningbo Chinese" *Proceedings of the 15<sup>th</sup> International Congress on Phonetic Sciences, Barcelona, August 3-9, 2003*, 3017-3020.
- Hurch, Bernhard. 1998. "Optimalität und Natürlichkeit." *ZAS Papers in Linguistics* 13, 115-139.
- Hyman, Larry. 2001. "The Limits of Phonetic Determinism in Phonology." In: Elizabeth Hume & Keith Johnson (eds.), *The Role of Speech Perception in Phonology*. San Diego and London: Academic Press, 141-185.
- Iivonen, Antti. 1984. "Die hochdeutschen und die schweizerhochdeutschen betonten Monophthonge." *Proceedings of the 10<sup>th</sup> International Congress of Phonetic Sciences, Dordrecht/Cinnaminson: Foris*, 191-196.
- Iivonen, Antti. 1987a. "Regional Differences in the Realization of Standard German Vowels." *Proceedings of the 11<sup>th</sup> International Congress of Phonetic Sciences, Tallinn: Academy of Sciences of the Estonian S.S.R., Vol. 4*, 161-164.
- Iivonen, Antti. 1987b. "Monophthonge des Gehobenen Wiener Deutsch." *Folia Linguistica, Tomus XXI/2-4*, 293-336.
- Jackson, Michael. 1988. "Phonetic Theory and Cross-linguistic Variation in Vowel Articula-

- tion." *UCLA Working Papers in Phonetics* 71.
- Jacobson, Leon C. 1978. „DhoLuo vowel harmony: a phonetic investigation.“ *UCLA Working Papers in Phonetics* 43.
- Jacobson, Leon C. 1980. "Voice-quality harmony in Western Nilotic languages." In: Robert M Vago (ed.), *Issues in Vowel Harmony*. Amsterdam: John Benjamins, 183-200.
- Jakobson, Roman & Morris Halle. 1961/2002<sup>3</sup>. "Tenseness and Laxness." Reprinted in: Roman Jakobson, *Selected Writings I, Phonological Studies*. Berlin, New York: de Gruyter, 550-555.
- Jenkins, James J., Winifred Strange & Sonja A. Trent. 1999. "Context-independent dynamic information for the perception of coarticulated vowels." *Journal of the Acoustical Society of America* 106 (1), 438-448.
- Jensen, Christian. 2001. "Schwa-assimilation in Danish Synthetic Speech." *Proceedings of the 7<sup>th</sup> European Conference on Speech Communication and Technology*. Eurospeech 2001. Scandinavia, 341-344.
- Jindra, Petr, Miroslav Eber & Josef Pešák. 2002. "The Spectral Analysis of Syllables in Patients using Dentures." *Biomed. Papers* 146 (2), 91-94.
- Johnson, Keith. 1997. *Acoustic and Auditory Phonetics*. Oxford: Blackwell.
- Joos; Martin. 1948. "Acoustic Phonetics." *Language* 24 (2), Suppl. (= *Language Monograph* Nr. 23).
- Jørgensen, Hans Peter. 1969. "Die gespannten und ungespannten Vokale in der norddeutschen Hochsprache mit einer spezifischen Untersuchung der Struktur ihrer Formantfrequenzen." *Phonetica* 19, 217-245.
- Jessen, Michael. 1993. "Stress conditions on vowel quality and quantity in German." *Working Papers of the Cornell Phonetics Laboratory* 8, 1-27.
- Jessen, Michael. 1997. "Speaker-specific information in voice quality parameters." *The International Journal of Speech, Language and the Law* 4 (1), 84-103.
- Jessen, Michael, Krzysztof Marasek, Katrin Schneider & Kathrin Clahßen. 1995. "Acoustic Correlates of Word Stress and the tense/lax Opposition in the Vowel System of German." *Proceedings of the 13<sup>th</sup> International Congress on Phonetic Sciences*. Stockholm. Vol 4, 428-431.
- Johnson, Keith & Jack Martin. 2001. "Acoustic Vowel Reduction in Creek: Effects of Distinctive Length and Position in the Word." *Phonetica* 58, 81-102.
- Jones, Jeffery A. & Kevin G. Munhall. 2003. "Learning to produce speech with an altered vocal tract: The role of auditory feedback." *Journal of the Acoustical Society of America* 113 (1), 532-543.
- Jongman, Allard, Marios Fourakis & Joan A. Sereno. 1989. "The Acoustic Vowel Space of Modern Greek and German." *Language and Speech* 32 (3), 221-248.
- Keating, Patricia A. 1990. "The window model of coarticulation: articulatory evidence." In: John Kingston & Mary E. Beckman (eds.), *Papers in Laboratory Phonology I. Between Grammar and Physics of Speech*. Cambridge: Cambridge University Press, 451-470.
- Kent, Ray D. 1983. "The Segmental Organization of Speech." In: Peter F. MacNeilage (ed.), *The Production of Speech*. New York, Heidelberg, Berlin: Springer Verlag, 57-89.
- Kingston, John & Randy Diehl. 1994. "Phonetic knowledge." *Language* 71, 419-454.
- Kleber, Felicitas & Nadine Klippfahn. 2006. "An acoustic investigation of secondary stress in German." *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel*, AIPUK 37, 1-18.
- Kloeke, Wus van Lessen. 1982. *Deutsche Phonologie und Morphologie: Merkmale und*

- Markiertheit. Tübingen: Niemeyer.
- Kochanski, Greg, Esther Grabe, John Coleman & Burton Rosner. 2005. "Loudness predicts prominence: Fundamental frequency lends little." *Journal of the Acoustical Society of America* 118 (2), 1038-1054.
- Kohler, Klaus J. 1995<sup>2</sup>. Einführung in die Phonetik des Deutschen. Berlin: Erich Schmidt Verlag.
- Kohler, Klaus J. 1999. "German." *Handbook of the International Phonetic Association*. Cambridge: Cambridge University Press, 86-89.
- Kohler, Klaus J. 2000. "Investigating Unscripted Speech: Implications for Phonetics and Phonology." *Phonetica* 57, 85-94.
- Kouznetsov, V. 2001. "Spectral Dynamics and Classification of Russian Vowels." *Proceedings of the 11<sup>th</sup> Session of the Russian Acoustical Society*, 439-442.
- Kozasa, Tomoko. 2002. "Duration and F0 cues for vowel length in Japanese." *Journal of the Acoustical Society of America* 111 (5), 2365.
- Kozhevnikov, Valery A. & L. A. Chistovich. 1965. *Speech: Articulation and Perception*. Washington, D. C.: Joint Publications Research Service.
- Kuehn, David P. & Kenneth L. Moll. 1976. "A cineradiographic study of VC and CV articulatory velocities." *Journal of Phonetics* 4, 303-320.
- Kühnert, Barbara & Francis Nolan. 1999. "The origin of coarticulation." In: William J. Hardcastle & Nigel Hewlett (eds.), *Coarticulation. Theory, Data and Techniques*. Cambridge: Cambridge University Press, 7-31.
- Künzel, Hermann. 2001. "Beware of the 'telephone effect': the influence of telephone transmission on the measurements of formant frequencies." *The International Journal of Speech, Language and the Law* 8 (1), 80-99.
- Künzel, Hermann, Angelika Braun & Ulrich Eysholdt. 1992. *Einfluß von Alkohol auf Sprache und Stimme*. Heidelberg: Kriminalistik Verlag.
- Ladefoged, Peter. 1964. *A Phonetic Study of West African Languages. An Auditory-instrumental Survey*. Cambridge: Cambridge University Press.
- Ladefoged, Peter. 1982<sup>2</sup>. *A Course in Phonetics*. New York etc.: Harcourt Brace Jovanovich, Inc.
- Ladefoged, Peter. 2001. *Vowels and Consoants*. Oxford: Blackwell.
- Ladefoged, Peter, Joseph DeClerk, Mona Lindau & George Papçun. 1972. „An auditory motor theory of speech production.“ *UCLA Working Papers in Phonetics* 22, 48-75.
- Ladefoged, Peter, Richard Harshman, Louis Goldstein & Lloyd Rice. 1978. "Generating vocal tract shapes from formant frequencies." *Journal of the Acoustical Society of America* 64 (4), 1027-1035.
- Ladefoged, Peter & Ian Maddieson. 1996. *The Sounds of the World's Languages*. Oxford, Cambridge: Blackwell Publishers.
- Lahiri, Aditi, Letitia Gewirth & Shiela E. Blumstein. 1984. "A Reconsideration of Acoustic Invariance for Place of Articulation in Diffuse Stop Consonants: Evidence from a Cross-language Study." *Journal of the Acoustical Society of America* 76 (2), 391-404.
- Lehiste, Ilse & Gordon E. Peterson. 1961. "Some basic considerations in the analysis of intonation." *Journal of the Acoustical Society of America* 33, 419-425.
- Liljencrants J. & Björn Lindblom. 1972. "Numerical simulation of vowel quality systems: The role of perceptual contrast." *Language* 48, 839-862.
- Lin, Qiguang & Gunnar Fant. 1989. "Vocal-tract Area-Function Parameters from Formant Frequencies." *Proceedings of the European Conference on speech Communication and*

- Technology. Eurospeech 89*. Paris. Vol 2, 673-676.
- Lindau, Mona. 1979. "The Feature Expanded." *Journal of Phonetics* 7, 163-176.
- Lindau, Mona. 1987. "Tongue mechanisms in Akan and Luo." *UCLA Working Papers in Phonetics* 68, 46-57.
- Lindblom, Björn. 1963. "Spectrographic study of vowel reduction." *Journal of the Acoustical Society of America* 35, 1773-1781.
- Lindblom, Björn. 1983. "Economy of Speech Gestures." In: Peter F. MacNeilage (ed.), *The Production of Speech*. New York, Heidelberg, Berlin: Springer Verlag, 217-245.
- Lindblom, Björn. 1986. "Phonetic universals in vowel systems." In: John J. Ohala & Jeri J. Jaeger (eds.), *Experimental Phonology*. Orlando, FL: Academic Press, 13-44.
- Lindblom, Björn. 1990. "Explaining Phonetic Variation. A Sketch of the H and H Theory." In: William J. Hardcastle & Alain Marchal (eds.), *Speech Production and Modelling*. Dordrecht, Boston, London: Kluwer Academic Publishers, 403-439.
- Lindblom, Björn. 2003. "Patterns of Phonetic Contrast: Towards a unified explanatory framework" *Proceedings of the 15<sup>th</sup> International Congress on Phonetic Sciences*, Barcelona, August 3-9, 2003, 39-42.
- Lindblom, Björn. 2004. "The organization of speech movements: Specification of units and modes of control." *From Sound to Sense: 50+ years of discoveries in speech communication*, Conference Proceedings (CD publication), Research Laboratory of Electronics, M.I.T., Cambridge, MA, USA, B - 86-97.
- Lindblom, Björn & Michael Studdert-Kennedy 1967. "On the Role of Formant Transitions in Vowel Recognition." *Journal of the Acoustical Society of America* 42 (4), 830-843.
- Lindblom, Björn, Harvey M. Sussman, Golnaz Modarresi & Elizabeth Burlingame. 2002. "The Trough Effect: Implications for Speech Motor Programming." *Phonetica* 59, 245-262.
- Lindblom, Björn & Harvey M. Sussman. 2004. "Principal components analysis of tongue shapes in symmetrical VCV utterances." *TMH-QPSR* 44, 1-4.
- Lobanov, Boris M. 1971. "Classification of Russian Vowels Spoken by Different Speakers." *Journal of the Acoustical Society of America* 49 (2), 606-608.
- Low, Ee Ling & Esther Grabe. 1995. "Prosodic patterns in Singapore English." *Proceedings of the 13<sup>th</sup> International Congress on Phonetic Sciences*. Stockholm. Vol 3, 636-639.
- Low, Ee Ling, Esther Grabe & Francis Nolan. 2000. "Quantitative characterizations of speech rhythm: Syllable-Timing in Singapore English." *Language and Speech* 43 (4), 377-401.
- Löfqvist, Anders. 1986. "Comment on James H. Abbs." In: Joseph S. Perkell & Dennis H. Klatt (ed.), *Invariance and Variability in Speech Processes*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 219-220.
- Löfqvist, Anders. 1997. "Theories and Models of Speech Production." In: William J. Hardcastle & John Laver (eds.), *The Handbook of Phonetic Sciences*. Oxford and Cambridge: Blackwell, 405-426.
- Löfqvist, Anders & Vincent L. Gracco. 2002. "Control of oral closure in lingual stop consonant production." *Journal of the Acoustical Society of America* 111 (6), 2811-2827.
- Lubker, James F. 1981. "Temporal Aspects of Speech Production: Anticipatory Labial Coarticulation." *Phonetica* 38, 51-65.
- Lubker, James F. & Thomas Gay. 1982. "Anticipatory labial coarticulation: Experimental, biological, and linguistic variables." *Journal of the Acoustical Society of America* 71 (2), 437-448.
- Maddieson, Ian. 1984. *Patterns of sounds*. Cambridge: Cambridge University Press.
- Maddieson, Ian. 1997. "Phonetic Universals." In: William J. Hardcastle & John Laver (eds.),

- The Handbook of Phonetic Sciences. Oxford and Cambridge: Blackwell, 619-639.
- Madelska, Liliana & Wolfgang U. Dressler. 1996. "Postlexical stress processes and their segmental consequences illustrated in Polish and Czech." In: Bernhard Hurch & Richard A. Rhodes (eds.), *Natural Phonology: The State of the Art*. Berlin, New York: Mouton de Gruyter, 189-200.
- Magen, Harriet S. 1997. "The extent of vowel-to-vowel coarticulation in English." *Journal of Phonetics* 25, 187-205.
- Maeda, Shinji. 1991. "On articulatory and acoustic variabilities." *Journal of Phonetics* 19, 321-331.
- Maeda, Shinji. 1999. "Labialization during /k/ followed by a rounded vowel is not anticipation but the auditorily required articulation." *Proceedings of the 14<sup>th</sup> International Congress on Phonetic Sciences, San Francisco, August 1-7, 1999*, 41-44.
- Manuel, Sharon Y. 1990. "The role of contrast in limiting vowel-to-vowel coarticulation in different languages." *Journal of the Acoustical Society of America* 88 (3), 1286-1298.
- Manuel, Sharon Y. & Kenneth N. Stevens. 1995. "Formant Transitions: Teasing apart Consonant and Vowel Contributions." *Proceedings of the 13<sup>th</sup> International Congress on Phonetic Sciences*. Stockholm. Vol 4, 436-439.
- Markel John D. & Augustine H. Gray. 1976. *Linear Prediction of Speech*. Berlin: Springer.
- Martin, Pierre 1998. "À Québec, a-t-on l'schwa?" In: Yves Duhoux (ed.), *Lingue et langues. Hommage à Albert Maniet*. Leuven: Peeters, 163-179.
- McAllister, Robert & Olle Engstrand. 1991. "Some cross language aspects of co-articulation." *PERILUS XIV*, 7-10.
- McGowan, Richard S. 2004. "Secant lines in a preliminary study of phonetic reduction of /i/ and /ε/." *Journal of the Acoustical Society of America* 116 (4), 2324-2337.
- McRobbie-Utasi, Zita & Donna Starks. 2001. „The KIT Vowel in New Zealand English: Implications of Variation in the Speech of an Interviewer in Rapid Surveys.“ *Proceedings of the 2001 Annual Conference of the Canadian Linguistic Association*, 145-157.
- Mengel, Andreas. 1997. "Das akustische Korrelat des deutschen Wortakzents." Vortrag gehalten anlässlich der Konferenz für Elektronische Sprachsignalverarbeitung (ESSV 1997) in Cottbus, <http://www.andreasmengel.de/pubs/essv-cottbus-97.html> .
- Mengel, Andreas. 2000. *Deutscher Wortakzent. Symbole, Signale*. Norderstedt: Books on Demand GmbH.
- Menzerath, Paul. 1935. *Der Stand der heutigen Lautwissenschaft. Betrachtungen zum II. Kongreß für Phonetik und Phonologie (London 22.-26. Juli 1935)*. Debrecen: Wärbel.
- Menzerath, Paul & Armando de Lacerda. 1933. *Koartikulation, Steuerung und Lautabgrenzung*. Berlin und Bonn: Ferd. Dümmers Verlag.
- Miller, James D. 1989. "Auditory-perceptual interpretation of the vowel." *Journal of the Acoustical Society of America* 85, 2114-2134.
- Möhler, Gregor & Grzegorz Dogil. 1995. "Test environment for the two level model of Germanic prominence." *Proceedings of the 4<sup>th</sup> European Conference on Speech Communication and Technology. Eurospeech 1995*. Madrid, Vol. 2, 1019-1022.
- Moll, Kenneth L. & Raymond G. Daniloff. 1971. "Investigation of the Timing of Velar Movements during Speech." *Journal of the Acoustical Society of America* 50 (2), 678-684.
- Moon, Seung-Jae & Björn Lindblom 1994. "Interaction between duration, context and speaking style in English stressed vowels." *Journal of the Acoustical Society of America* 96 (1), 40-55.
- Mooshammer, Christine, Philip Hoole & Barbara Kühnert. 1995. "On loops." *Journal of*

- Phonetics 23, 3-21.
- Mooshammer, Christine. 1998. "Experimentalphonetische Untersuchungen zur artikulatorischen Modellierung der Gespanntheitsopposition im Deutschen." *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München* 36, 3-192.
- Moosmüller, Sylvia. 1987. *Soziophonologische Variation im gegenwärtigen Wiener Deutsch. Eine empirische Untersuchung*. Stuttgart: Steiner (= *Zeitschrift für Dialektologie und Linguistik* 56).
- Moosmüller, Sylvia. 1991. *Hochsprache und Dialekt in Österreich*. Wien: Böhlau.
- Moosmüller, Sylvia. 1997a. "Phonological variation in speaker identification." *Forensic Linguistics* 4 (1), 29-47.
- Moosmüller, Sylvia. 1997b. "Diphthongs and the process of Monophthongization in Austrian German: A First Approach." *Proceedings of the 5<sup>th</sup> European Conference on Speech Communication and Technology*. Eurospeech 1997. Rhodes, 787-790.
- Moosmüller, Sylvia. 1997c. "Regionalspezifische Realisierung von Diphthongen und ihre Auswirkung auf die Monophthongierung." *Grazer Linguistische Studien* 47, 83-98.
- Moosmüller, Sylvia. 1997d. "Die Auswirkung von Aktivierung auf phonologische Entdeutlichungsprozesse." In: Foltin, Robert & Wolfgang U. Dressler (eds.), *Phonologie und Psychophysiologie*. Wien: Österreichische Akademie der Wissenschaften, 23-38.
- Moosmüller, Sylvia. 1999. "The modulation of formant amplitude during assimilation processes of homorganic and quasi-homorganic V1##V2 sequences." *Proceedings of the 14<sup>th</sup> International Congress on Phonetic Sciences*, San Francisco, August 1-7, 1999, 1329-1332.
- Moosmüller, Sylvia. 2002. "Der Stellenwert der phonologischen und phonetischen Variation in der Sprechererkennung." In: Braun, Angelika & Herbert Masthoff (eds.), *Phonetics and its applications: Festschrift for Jens-Peter Köster on the Occasion of his 60<sup>th</sup> Birthday*. Stuttgart: Steiner (= *Zeitschrift für Dialektologie und Linguistik*, Beihefte H. 121), 97 – 109.
- Moosmüller, Sylvia. 2006. "Some relevant aspects of vowel formant interpretation." *Grazer Linguistische Studien* 65, 53-73.
- Moosmüller, Sylvia. 2007a. "When does lip protrusion start in Standard Austrian German? An acoustic investigation." *Proceedings of the 16<sup>th</sup> International Congress on Phonetic Sciences*, Saarbrücken, August 6-10, 2007, to appear.
- Moosmüller, Sylvia. 2007b. "Phonetics needs Phonology." In: *Festschrift für Rodmonga Potapova*. (to appear).
- Moosmüller, Sylvia & Ralf Vollmann. 2000. "The spread of the Viennese monophthongization: A socio-phonetic analysis." In: Schaner-Wolles, Chris, John R. Rennison & Friedrich Neubarth (eds.), *Naturally! Linguistic studies in honour of Wolfgang Ulrich Dressler presented on the occasion of his 60th birthday*. Torino: Rosenberg & Sellier, 327-335.
- Moosmüller, Sylvia & Ralf Vollmann. 2001. "„Natürliches Driften“ im Lautwandel: Die Monophthongierung im Österreichischen Deutsch." *Zeitschrift für Sprachwissenschaft* 20/I, 42-65.
- Moosmüller, Sylvia & Theodor Granser. 2003. "The Vowels of Standard Albanian." *Proceedings of the 15<sup>th</sup> International Congress on Phonetic Sciences*, Barcelona, August 3-9, 2003, 659-662.
- Moosmüller, Sylvia & Catherine O. Ringen. 2004. "Voice and Aspiration in Austrian German Plosives." *Folia Linguistica* XXXVIII/1-2, 43 – 62.
- Moosmüller, Sylvia & Theodor Granser. 2006. "The spread of Standard Albanian: An illustration based on an analysis of vowels." *Language, Variation and Change* 18 (2), 121-

140.

- Moulton, William G. 1962. *The Sounds of English and German. A systematic analysis of the contrasts between the sound systems.* Chicago and London: University of Chicago Press.
- Mrayati, M., René Carré & Bernard Guerin. 1988. "Distinctive Regions and Modes: A new Theory of Speech Production." *Speech Communication* 7, 257-286.
- Nakamura, Mitsuhiro. 2005. "Parametric Phonetics: an exercise in the dynamic characterisation of sound patterns." *Phonetic Teachings and Learning Conference, London, Department of Phonetics and Linguistics.* CD-Rom.
- Narayanan, S.S., A. A. Alwan, & K. Haker. 1995. "An articulatory study of fricative consonants using magnetic resonance imaging." *Journal of the Acoustical Society of America* 98 (3), 1325-1347.
- Nearey, Terrance M. 1989. "Static, Dynamic, and Relational Properties in Vowel Perception." *Journal of the Acoustical Society of America* 85 (5), 2088-2113.
- Nearey, Terrance M. & Peter F. Assmann. 1986. "Modeling the Role of Inherent Spectral Change in Vowel Identification." *Journal of the Acoustical Society of America* 80 (5), 1297-1308.
- Neweklowsky, Gerhard. 1975. "Spezifische Dauer und spezifische Tonhöhe der Vokale." *Phonetica* 32, 38-60.
- Nolan, Francis. 2002. "The 'telephone effect' on formants: a response." *The International Journal of Speech, Language and the Law* 9 (1), 74-82.
- Nolan, Francis, Tara Holst & Barbara Kühnert. 1996. "Modelling [s] to [ʃ] accommodation in English." *Journal of Phonetics* 24 (1), 113-137.
- Nolan, Francis & Catalin Grigoras. 2005. "A case for formant analysis in forensic speaker identification." *The International Journal of Speech, Language and the Law* 12 (2), 143-173.
- Nord, Lennart. 1974. "Vowel Reduction—Centralization or Contextual Assimilation?" In: Gunnar Fant (ed.), *Speech Communication 2. Speech Production and Synthesis by Rules.* Stockholm: Almqvist & Wiksell International, 149-154.
- Nord, Lennart. 1986. "Acoustic studies of vowel reduction in Swedish." *STL-QPSR* 4/1986, 19-36.
- Nowak, Pawel Marcin. 2006. *Vowel Reduction in Polish.* PhD, University of California, Berkeley.
- Oh, Eunjin. 2002. "Fronting of back vowels in coronal contexts: a cross-language study." *Studies in Phonetics, Phonology and Morphology* 8 (2), 239 – 254.
- Ohala, John J. 1983. "The origin of sound patterns in vocal tract constraints." In: Peter F. MacNeilage (ed.), *The Production of Speech.* New York, Heidelberg, Berlin: Springer Verlag, 189-216.
- Ohala, John 1993. "Coarticulation and Phonology." *Language and Speech* 36, 155-170.
- Öhmann, Sven E. G. 1966. "Coarticulation in VCV utterances: Spectrographic Measurements." *Journal of the Acoustical Society of America* 39 (1), 151-168.
- Ostry, David J., Paul L. Gribble & Vincent L. Gracco. 1996. "Coarticulation of Jaw Movements in Speech Production: Is Context Sensitivity in Speech Kinematics Centrally Planned?" *The Journal of Neuroscience* 16 (4), 1570-1579.
- Ostry, David J. & Vincent L. Gracco. 1995. "Is intra-articulator speech coarticulation planned?" *Journal of the Acoustical Society of America* 97 (5), 3400 (A).
- Padgett, Jaye & Marija Tabain. 2005. "Adaptive Dispersion Theory and Phonological Vowel Reduction in Russian." *Phonetica* 62 (1), 14-54.

- Parker, Steve. 2001. "The Acoustic Qualities of Bora Vowels." *Phonetica* 58, 179-195.
- Perkell, Joseph S. 1980. "Phonetic Features and the Physiology of Speech Production." In: Brian Butterworth (ed.), *Language Production. Volume 1: Speech and Talk*. London, New York: Academic Press, 337-372.
- Perkell, Joseph S. 1997. "Articulatory Processes." In: William J. Hardcastle & John Laver (eds.), *The Handbook of Phonetic Sciences*. Oxford and Cambridge: Blackwell, 333-379.
- Perkell, Joseph S. & Melanie L. Matthies. 1992. "Temporal measures of anticipatory labial coarticulation for the vowel /u/: Within- and cross-subject variability." *Journal of the Acoustical Society of America* 91 (5), 2911-2925.
- Perkell, Joseph S., Majid Zandipour, Melanie L. Matthies & Harlan Lane. 2002. "Economy of effort in different speaking conditions. I. A preliminary study of intersubject differences and modelling issues." *Journal of the Acoustical Society of America* 112, 1627-1641.
- Perkell, Joseph S., Frank H. Guenther, Harlan Lane, Melanie L. Matthies, Ellen Stockmann, Mark Tiede & Majid Zandipour. 2004. "The distinctness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts." *Journal of the Acoustical Society of America* 116, 2338-2344.
- Perkell, Joseph S., Frank H. Guenther, Harlan Lane, Nicole Marrone, Melanie L. Matthies, Ellen Stockmann, Mark Tiede & Majid Zandipour. 2006. "Production and Perception of Phoneme Contrasts Covary Across Speakers." In: Jonathan Harrington & Marija Tabain (eds.), *Speech Production. Models, Phonetic Processes, and Techniques*. New York and Hove: Psychology Press, 69-84.
- Perrier, Pascal. 2005. "Control and representations in speech production." *ZAS Papers in Linguistics* 40, 109-132.
- Perrier, Pascal, Yohan Payan, Majid Zandipour & Joseph Perkell. 2003. "Influences of tongue biomechanics on speech movements during the production of velar stop consonants: A modeling study." *Journal of the Acoustical Society of America* 114 (3), 1582-1599.
- Peterson, Gordon E. & Harold L. Barney. 1952. "Control methods used in a study of vowels." *Journal of the Acoustical Society of America* 24, 175-184.
- Pétursson, Magnus. 1992<sup>3</sup>. *Lehrbuch der isländischen Sprache*. Hamburg: Buske.
- Pfitzinger, Hartmut R. 2005. "Towards functional modelling of relationships between the acoustics and perception of vowels." *ZAS Papers in Linguistics* 40, 133-144.
- Piterman, Michel. 2000. "Effect of speaking rate and contrastive stress on formant dynamics and vowel perception." *Journal of the Acoustical Society of America* 107, 3425-3437.
- Podevska, Robert J. & Niken Adisasmito-Smith. 1999. "Acoustic Investigation of the Vowel Systems of Buginese and Toba Batak." *Proceedings of the 14<sup>th</sup> International Congress on Phonetic Sciences, San Francisco, August 1-7, 1999*, 535-538.
- Pols, Louis W. C., L. Y. Th. Van der Kamp & R. Plomp. 1969. „Perceptual and Physical Space of Vowel Sounds.“ *Journal of the Acoustical Society of America* 46 (2), 458-467.
- Port, Robert F. to appear. "How are words stored in memory? Beyond phones and phonemes." *New Ideas in Psychology*.
- Poupplier, Marianne, Adam Buchwald & Maureen Stone. 2004. „A Tagged Cine-MRI and Ultrasound Study of the German Vowel System.“ *Working Papers and Technical Reports, No 2004 (6)*, Vocal Tract Visualization Laboratory, University of Maryland Dental School, 18-55.
- Pulleyblank, E.G. 2003. "Non-contrastive Features or Enhancement by Redundant Features?" *Language and Linguistics* 4.4, 713-755.
- Ramers, Karl Heinz. 1988. *Vokalquantität und –qualität im Deutschen*. Tübingen: Niemeyer.

- Ramus, Franck, Marina Nespor & Jacques Mehler. 1999. "Correlates of linguistic rhythm in the speech signal". *Cognition* 73 (3), 265-292.
- Ramus, Franck. 2002. "Acoustic correlates of linguistic rhythm: Perspectives." In: Bel, Bernard & Isabelle Marlien (eds.), *Proceedings of Speech Prosody 2002*, 115-120.
- Raphael, Lawrence J. & Fredericka Bell-Berti. 1975. "Tongue musculature and the feature of tension in English vowels." *Phonetica* 32, 61-73.
- Rausch, Arsen. 1972. "Untersuchungen zur Vokalartikulation im Deutschen." In: Heinrich Kelz & Arsen Rausch (eds.), *Beiträge zur Phonetik*. Hamburg: Helmut Buske Verlag, 35-82.
- Recasens, Daniel. 1999. "Lingual coarticulation." In: William J. Hardcastle & Nigel Hewlett (eds.), *Coarticulation. Theory, Data and Techniques*. Cambridge: Cambridge University Press, 81-104.
- Recasens, Daniel, Maria Dolores Pallarès & Jordi Fontdevila. 1997. "A model of lingual coarticulation based on articulatory constraints." *Journal of the Acoustical Society of America* 102 (1), 544-561.
- Reiffenstein, Ingo. 1983. "Deutsch in Österreich." In: Ingo Reiffenstein (ed.), *Tendenzen, Formen und Strukturen der deutschen Standardsprache nach 1945*. Marburg/Lahn: Elwert, 15-27.
- Restle, David. 1998. *Silbenschnitt – Quantität – Kopplung: Zur Geschichte, Charakterisierung und Repräsentation der Anschlußprosodie unter dem Blickwinkel einer Oszyllationssilbentheorie*. Doctoral Dissertation, University of Munich.
- Restle, David. 1999. "Reduplication as a pure constituent copying." In: John Rennison & Klaus Kühnhammer (eds.), *Phonologica 1996*. The Hague: Thesus, 195-217.
- Ronneberger-Sibold, Elke. 1999. "Ambisyllabic consonants in German: Evidence from dialectal pronunciation of lexical creations." In: John Rennison & Klaus Kühnhammer (eds.), *Phonologica 1996*. The Hague: Thesus, 247-271.
- Rouas, Jean-Luc, Jérôme Farinas, François Pellegrino & Régine André-Obrecht. 2005. "Rhythmic unit extraction and modelling for automatic language identification." *Speech Communication* 47, 436-456.
- Roy, Johanna-Pascal, Rudolph Sock, Béatrice Vaxelaire & Fabrice Hirsch. 2003. "Auditory effects of anticipatory and carryover coarticulation." *Proceedings of the 6<sup>th</sup> International Seminar on Speech Production*, Macquarie University, Sydney, December 7-10, 2003.
- Savariaux, Christophe, Pascal Perrier & Jean-Pierre Orliaguet. 1995. "Compensation strategies for the perturbation of the rounded vowel [u] using a lip-tube: a study of the control space in speech production." *Journal of the Acoustical Society of America* 98 (5), 2428-2442.
- Sawusch, James R. 1996. "Effects of duration and formant movement on vowel perception." 4<sup>th</sup> International Conference on Spoken Language Processing ICSLP-1996, 2482-2485.
- Scarborough, Rebecca Anne. 2004. "Lexical Confusability and Degree of Coarticulation." *Proceedings of the 29<sup>th</sup> Meeting of the Berkeley Linguistics Society*, February 14-17, 2003.
- Schaeffler, Felix & Pär Wretling. 2002. "Long and short /a/ in Northern Swedish dialects." *TMH-QPSR* 44, 49-52.
- Schaeffler, Felix, Pär Wretling & Eva Strangert. 2002. "On the Development of a Quantity Typology for Swedish." In: Bel, Bernard & Isabelle Marlien (eds.), *Proceedings of Speech Prosody 2002*, 627-630.
- Schreuder, Maartje. 2006. *Prosodic Processes in Language and Music*. PhD Dissertation, Groningen Dissertations in Linguistics (GRODIL) 60, University of Groningen.
- Schröder, Konrad. 2001. "Englisch als Nachbarsprache, Englisch als internationale Sprache." In: Konrad Ehlich, Jakob Ossner & Harro Stammerjohann (eds.), *Hochsprachen in Europa*.

- Entstehung, Geltung, Zukunft. Freiburg im Breisgau: Fillibach. 301-308.
- Schwartz, Jean-Luc, Louis-Jean Boë, N. Vallée & Christian Abry. 1997. "Major trends in vowel system inventories." *Journal of Phonetics* 25 (3), 233-253.
- Sendlmeier, Wolfgang F. 1981. "Der Einfluß von Qualität und Quantität auf die Perzeption betonter Vokale des Deutschen." *Phonetica* 38, 291-308.
- Sievers, E. 1901. *Grundzüge der Phonetik zur Einführung in das Studium der Lautlehre der indogermanischen Sprachen*. 5. Auflage. Leipzig: Breitkopf & Härtel.
- Simpson, Adrian P. 1998. "Deutsche Vokale – Spontan und Gelesen." *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung (AIPUK)* 33, 33-217
- Simpson, Adrian P. 1999. "Fundamental Problems in Comparative Phonetics and Phonology: Does UPSID help to solve them?" *Proceedings of the 14<sup>th</sup> International Congress on Phonetic Sciences, San Francisco, August 1-7, 1999*, 349-352.
- Slifka, Janet. 2003. "Tense/lax vowel classification using dynamic spectral cues." *Proceedings of the 15<sup>th</sup> International Congress on Phonetic Sciences, Barcelona, August 3-9, 2003*, 33-38.
- Solé, Maria-Josep & John J. Ohala. 1991. "Differentiating between phonetic and phonological processes." *Proceedings of the 12<sup>th</sup> International Congress of Phonetic Sciences, Aix en Provence: 1991, Vol. 2*, 921-924.
- Stampe, David. 1979. *A Dissertation on Natural Phonology*. New York and London: Garland Publishing, Inc.
- Stevens, Kenneth N. 1972. "The quantal nature of speech: evidence from articulatory-acoustic data." In: Edward E. David Jr. & Peter B. Denes (ed.), *Human Communication: A Unified View*. New York: McGraw Hill, 51-66.
- Stevens, Kenneth N. 1989. "On the quantal nature of speech." *Journal of Phonetics* 17, 3-46.
- Stevens, Kenneth N. 1999. *Acoustic Phonetics*. Cambridge Mass.: The MIT Press.
- Stevens, Kenneth N. 2003. "Acoustic and Perceptual Evidence for Universal Phonological Features." *Proceedings of the 15<sup>th</sup> International Congress on Phonetic Sciences, Barcelona, August 3-9, 2003*, 33-38.
- Stevens, Kenneth N. & Arthur S. House. 1963. "Perturbation of vowel articulations by consonantal context: An acoustical study." *Journal of Speech and Hearing Research* 6, 111-128.
- Stevens, Kenneth N. & Shiela E. Blumstein. 1978. "Invariant Cues for Place of Articulation in Stop Consonants." *Journal of the Acoustical Society of America* 64 (5), 1358-1368.
- Stevens, Kenneth N., Sharon Y. Manuel & Melanie Matthies. 1999. "Revisiting Place of Articulation Measures for Stop Consonants: Implications for Models of Consonant Production." *Proceedings of the 14<sup>th</sup> International Congress on Phonetic Sciences, San Francisco, August 1-7, 1999*, 1117-1120.
- Strange, Winifred & Ocke-Schwen Bohn. 1998. "Dynamic specification of coarticulated German vowels: Perceptual and acoustical studies." *Journal of the Acoustical Society of America* 104, 488-504.
- Stratka, Georges. 1978. "A propos du classement articulatoire des voyelles." In: Hans-Heinrich Wängler (ed.), *Festschrift für Otto von Essen*. Hamburg: Buske, 437-460 (= *Hamburger Phonetische Beiträge* 25).
- Tabain, Marija. 2001. Review of Hardcastle, William J. & Nigel Hewlett (eds.) (1999). *Coarticulation: Theory, Data and Techniques*, Cambridge University Press. In: *Linguist List* 12.490, February 22<sup>nd</sup>, 2001.
- Tabain, Marija. 2003. "Effects of prosodic boundary on /aC/ sequences: Acoustic results."

- Journal of the Acoustical Society of America 113 (1), 516-531.
- Tabain, Marija, Pascal Perrier, Christophe Savariaux & Richard Beare. 2004. "An articulatory prosody of /u/: Motor equivalence." Proceedings of the 10<sup>th</sup> Australian International Conference on Speech Science & Technology, Macquarie University, Sydney, December 8-10, 2004, 432-435.
- Tabain, Marija & Pascal Perrier. (2005). "Articulation and acoustics of /i/ in preboundary position in French." *Journal of Phonetics* 33, 77-100.
- Theron, Karin. 2003. Temporal aspects of speech production in bilingual speakers with neurogenic speech disorders. Doctoral Thesis, University of Pretoria.
- Tiede, Mark K. 1996. "An MRI-based study of pharyngeal volume contrasts in Akan and English." *Journal of Phonetics* 24, 399-421.
- Valaczkai, László. 1998. Atlas deutscher Sprachlaute. Instrumentalphonetische Untersuchung der Realisierung deutscher Phoneme als Sprechlaute. Wien: Edition Praesens.
- van Bergem Dick R. 1994. "A model of coarticulatory effects on the schwa." *Speech Communication* 14 (2), 143-162.
- Van der Merwe, Anita. 1997. "A theoretical framework for the characterization of pathologic speech sensorimotor control." In: Malcolm R. McNeil (ed.), *Clinical management of sensorimotor speech disorders*. New York: Thieme, 1-25.
- Vanecek, Erich & Wolfgang U. Dressler. 1977. "Untersuchungen zur Sprechsorgfalt als Aufmerksamkeitsindikator." *Studia Psychologica* 19, 105-118.
- van Son, Rob J. J. H. 1993. Spectro-Temporal Features of Vowel Segments. Academisch Proefschrift, Faculteit der Letteren, Universiteit van Amsterdam.
- van Son, Rob J. J. H. & Louis C. Pols. 1990. "Formant frequencies of Dutch vowels in a text, read at normal and fast rate." *Journal of the Acoustical Society of America* 88 (4), 1683-1693.
- van Son, Rob J. J. H. & Louis C. Pols. 1992. "Formant movements of Dutch vowels in a text, read at normal and fast rate." *Journal of the Acoustical Society of America* 92 (1), 121-127.
- van Son, Rob J. J. H. & Louis C. Pols. 2002. "Evidence for efficiency in vowel production." 7<sup>th</sup> International Conference on Spoken Language Processing ICSLP '02, Vol 1, 37-40.
- Vaxelaire, Béatrice, Rudolph Sock, Johanna-Pascale Roy, Aline Ascii & Véronique Hecker. 2003. "Audible and Inaudible Anticipatory Gestures in French." Proceedings of the 15<sup>th</sup> International Congress on Phonetic Sciences, Barcelona, August 3-9, 2003, 447-450.
- Vazquez-Alvarez, Yolanda. 2005. "The trough effect: Can we predict tongue lowering from acoustic data alone?" <http://ultrafest.arizona.edu/papers/1>.
- Vollmann, Ralf. 2005. Descriptions of Tibetan Ergativity. Habilitationsschrift Graz.
- Vollmann, Ralf & Sylvia Moosmüller. 1999. "The change of diphthongs in Standard Viennese German: The diphthong /ae/." Proceedings of the 14<sup>th</sup> International Congress on Phonetic Sciences, San Francisco, August 1-7, 1999, 345-348.
- Wagner, Petra S. 2002. Vorhersage und Wahrnehmung deutscher Betonungsmuster. Inaugural-Dissertation, Universität Bonn.
- Wesener, Thomas. 2001. "Non-sequential phenomena in German function words." *Journal of the International Phonetic Association* 31 (1), 17-27.
- Whalen, Douglas H. & Andrea G. Levitt. 1995. "The intrinsic pitch of vowels." *Journal of Phonetics* 23, 349-366.
- Whalen, Douglas H., Brian Gick, Masanobu Kumada & Kiyoshi Honda. 1998. "Cricothyroid activity in high and low vowels: Exploring the automaticity of intrinsic F0." *Journal of Phonetics* 27, 125-142.

- Whalen Douglas H., Brian Gick & Philip S. LeSourd. 1999. "Intrinsic F0 in Passamaquoddy vowels." Papers from the 30<sup>th</sup> Algonquian conference. Winnipeg: University of Manitoba, 417-428.
- Widera, Christina. 2002. "Vowel Reduction in German Read Speech." In: Eric Keller, Gérard Bailly, Alex Monaghan, Jacques Terken & Mark Huckvale (eds.), *Improvements in Speech Synthesis. COST 258: The Naturalness of Synthetic Speech*. Chichester: John Wiley & Sons, 186-195.
- Widera, Christina & Thomas Portele. 1999. "Levels of reduction for German tense vowels." *Proceedings of the 6<sup>th</sup> European Conference on Speech Communication and Technology. Eurospeech 1999*. Budapest, 1695-1698.
- Wiese, Richard. 1996. *The Phonology of German*. Oxford: Clarendon Press.
- Wood, Sidney. 1975a. "What is the difference between English and Swedish dental stops?" *Lund Working Papers* 10, 174-193.
- Wood, Sidney. 1975b. "Tense and Lax Vowels – Degree of Constriction or Pharyngeal Volume?" *Lund Working Papers* 11, 109-134.
- Wood, Sidney. 1979. "A radiographic analysis of constriction location for vowels." *Journal of Phonetics* 7, 25-43.
- Wood, Sidney. 1982. *X-Ray and Model Studies of Vowel Articulation*. Lund Working Papers 23.
- Wood, Sidney. 1986. "The Acoustical Significance of Tongue, Lip, and Larynx Maneuvers in Rounded Palatal Vowels." *Journal of the Acoustical Society of America* 80 (2), 391-401.
- Wood, Sidney. 1987. "The History of the Classical Vowel Articulation Model: A Reply to Catford and Fischer-Jørgensen." *Proceedings of the 11<sup>th</sup> International Congress on Phonetic Sciences, Tallinn, August 1-7, 1987, Vol. 2*, 53-56.
- Wood, Sidney. 1991a. "Vowel Gestures and Spectra: From Raw Data to Simulation and Applications." *Proceedings of the 12<sup>th</sup> International Congress of Phonetic Sciences, Aix en Provence: 1991, Vol. 1*, 215-219.
- Wood, Sidney. 1991b. „Vertical, monovocalic and other ‘impossible’ vowel systems: a review of the articulation of the Karbardian vowels.“ *Linguistica* 45, 49-70.
- Wood, Sidney. 1991c. "X-ray data on the temporal co-ordination of speech gestures." *Journal of Phonetics* 19, 281-292.
- Wood, Sidney. 1996. "Assimilation or coarticulation? Evidence from the temporal co-ordination of tongue gestures for the palatalization of Bulgarian alveolar stops." *Journal of Phonetics* 24, 139-164.
- Wood, Sidney. 1997. "A cinefluorographic study of the temporal organization of articulator gestures: Examples from Greenlandic." *Speech Communication* 22, 207-225.
- Wood, Sidney & T. Pettersson. 1988. Vowel reduction in Bulgarian: the phonetic data and model experiments. *Folia Linguistica, Tomus XXII*, 3 – 4, 239-262.
- Wouters, Johan. & Michael W. Macon. 2002. "Effects of prosodic factors on spectral dynamics. I. Analysis." *Journal of the Acoustical Society of America* 111 (1), 417-427.
- Wurzel, Wolfgang U. 1970. *Studien zur deutschen Lautstruktur*. Berlin: Akademie-Verlag.
- Wurzel, Wolfgang U. 1980. "Der deutsche Wortakzent. Fakten-Regeln-Prinzipien." *Zeitschrift für Germanistik* 3, 299-318.
- Zhang, Zhaoyan, Suzanne Boyce, Carol Espy-Wilson & Mark Tiede. 2003. "Acoustic Strategies for Production of American English „retroflex“ /r/." *Proceedings of the 15<sup>th</sup> International Congress on Phonetic Sciences, Barcelona, August 3-9, 2003*, 1125-1128.
- Zint-Dyhr, Ingeborg. 2001. "Der Hochsprachenbegriff in der dänischen Sprachkultur." In:

Sylvia Moosmüller

Konrad Ehlich, Jakob Ossner & Harro Stammerjohann (eds.), *Hochsprachen in Europa. Entstehung, Geltung, Zukunft*. Freiburg im Breisgau: Fillibach. 101-110.

Zonneveld, Wim, Mieke Trommelen, Michael Jessen, Curtis Rice, Gösta Bruce & Kristjan Árnason. 1999. "Wordstress in West-Germanic and North-Germanic languages." In: Harry van der Hulst (ed.), *Word Prosodic Systems in the Languages of Europe*. Berlin, New York: Mouton de Gruyter, 477-603.

## Appendix

- 01 Im Gasthaus ums Eck gibt es köstliches Bier.
- 02 Bei Radio Lippe wird täglich ein "Held des Alltags" gewählt.
- 03 Durch diese Luke kannst du sehen, wie sich der Kater in der Sonne reckelt.
- 04 Die Tiger jagten das Tier, bis es erschöpft in der Hecke steckenblieb.
- 05 Als Heike Herrn Huber sah, ist sie zu Tode erschrocken.
- 06 Schade, daß Dreck an diesem Sakko ist, sonst hätte es keinen Makel.
- 07 Kannst du bitte diesen Egel aus der Lake entfernen?
- 08 Sie versuchte vergeblich, ihren Sohn vor dieser Sekte zu bewahren.
- 09 Raube niemals eine Kassa aus.
- 10 An den Küsten streifen heimtückische Köter herum, die man besser an eine Kette legen sollte.
- 11 Dieses Kipfel, jetzt hat sie doch glatt vergessen, den Koffer zu packen.
- 12 Vergiß nicht, Kreis und Rechteck zu zeichnen.
- 13 Herr Klatt wird uns einige Lieder von Schubert singen.
- 14 Ich weiß, daß die Lage sehr ernst ist, dennoch wirst du deine Karriere doch nicht an den Nagel hängen.
- 15 Ich bitte dich, stecke doch den Rosenstock in das Beet.
- 16 Mit dem Rade wirst du bald Griechenland erreichen.
- 17 Leider wird er diesmal nichts backen.
- 18 Kader A1 für Junioren im Weitsprung ist noch nicht definitiv gebildet.
- 19 Hans klaubte Zwetschken und Äpfel vom Boden auf.
- 20 Die Fotos wird sie bald entwickeln lassen müssen, damit sie sie als Beweismittel heranziehen kann.
- 21 Der neue Mieter will mit Uta Sieke nicht Karten spielen.
- 22 Bei diesen wilden Jagden fand so mancher Gote den Tod.
- 23 Unsere liebe Ruth hat wieder einmal ein köstliches Essen zubereitet.
- 24 Ekel ergriff sie angesichts dieser riesigen Krake.
- 25 Aber die Krabbe und die Raupe, die ihr Sohn gefangen hatte, gefiel ihr auch nicht besser.
- 26 1954 wurde Jim's Pier in Texas eröffnet.
- 27 Es ist höchst unwahrscheinlich, daß dir eine Pause gegeben wird.
- 28 Wer kein "ü" sprechen kann, sagt Kir statt Kür.

- 29 Du wirst zugeben müssen, daß dieser Kopf ein ausgezeichnetes Modell für den Bildhauer ist.
- 30 Er suhlt sich in kühnen Deutungen über den von einer königlichen Sippe geführten Pazifistenstaat.
- 31 Sie unterrichtete lange Zeit am logopädisch-phoniatischen Institut.
- 32 In vielen Büchern wird über die Situation der Mägde in den Herrschaftshäusern berichtet.
- 33 Die Göttin der Liebe sah spöttisch auf die aus ihrer Sicht doofen Menschen herab.
- 34 Heinz regelt das für dich, er wird ein Lokal finden.
- 35 In einer Stadt südlich von Freistadt erhob sich eine rege Diskussion über die Bedeutung verschiedener Käfer.
- 36 Gieß doch bitte mal die Blumen, aber drück dabei nicht das Unkraut platt.
- 37 Kannst du bitte mit Gabel und Messer essen!
- 38 Die Milch geht zur Neige, wir haben nur noch einen Liter.
- 39 Unsere Ratte hat schon wieder das Kabel angeknabbert.
- 40 Sie probiert aus, ob diese subtilen Unterschiede relevant sind.
- 41 Mit Pflug und Egge wird die Aussaat vorbereitet, mit Egge, Walze und Hacke für gute Wachstumsbedingungen von Getreide oder Feldfrüchten gesorgt.
- 42 Der Apfelsaft wird immer trüber, je länger er steht.
- 43 Mit einer Leiter kannst du das Blatt vom Baume holen.
- 44 Im Zoo finden wir Kamele, Luchse, Füchse, Ottern, Widder, Vögel, Ziegen und viele andere Tiere.
- 45 Rate mal, wen ich gestern in der Oper getroffen habe.
- 46 Die APA berichtete über den Film "Die Häupter meiner Lieben".
- 47 Sein Sternzeichen ist Widder, daher fühlt er sich doppelt stark.
- 48 Da ist die Donau, da kommt er bestimmt nicht drüber.
- 49 Hast du schon neuen Mate Tee gekauft?
- 50 Als sie den Tisch deckte, vergaß sie fast, den Zucker in die Dose zu geben.
- 51 Der Ober erschrak über die tote Made im Salat.
- 52 Leider hab ich vergessen, die Arbeitsblätter des Buchs zu bestellen.
- 53 Herr Dr. Siepe wird deiner Krankheit zu Leibe rücken.
- 54 Dieser Weg, der in die Leopold Faust – Gasse führt, ist mit Kies bestreut.
- 55 Treck Informationen gibt es vierteljährlich in dieser Zeitschrift.
- 56 Die Verständigung der Staaten wird heute zu Grabe getragen, mokiert sich der Politiker.
- 57 Sie hätte sagen müssen, daß sie sich einen Zopf flechten möchte.
- 58 Ein weiterer Nahost-Gipfel wird nächstes Monat stattfinden.

- 59 Die Trauben müssen durch mehrere Siebe gedrückt werden, damit der Saft auch wirklich rein wird.
- 60 Mit Gier wirst du kaum bekommen, was du gerne kaufen möchtest.
- 61 Sie müssen das Mieder nicht bar bezahlen, es ist aus reiner Seide und daher sehr teuer.
- 62 Hierbei handelt es sich um eine unmögliche Form von musikalischer Darbietung.
- 63 Der zweitgrößte niederländische Pharmakonzern hat die Vitamin-Division von Roche übernommen.
- 64 Im Griechischen heißt "Greis" géron.
- 65 Das Paar hatte sich in Leipe, einer kleinen Stadt in Deutschland, niedergelassen.
- 66 Die Eisengießerei von Meuselwitz bietet Produkte aus einem Guß.
- 67 Gib doch bitte mal dem Schaf das Futter.
- 68 Zwei Dukaten vergaß der Prinz in dem Motel.
- 69 Im Servicecenter "Heim und Garten" können sie wunderschöne Gartenmöbel kaufen.
- 70 Zünd ja nicht das Haus an, während ich das Buch zur Post trage.
- 71 Heinz glaubte noch immer nicht, daß die Hupe des Puchs funktioniert.
- 72 Es kriechen winzige Maden im Speck.
- 73 Mit Kassaschluß müssen alle das Geschäft verlassen.
- 74 Im Hintergrund agiert der Sekretär des Finanzministers.
- 75 Hast du schon Köder für den Fischfang gekauft?
- 76 Der Herzog ritt nach Ried, wo er dem Pfarrer riet, sich aus der Politik rauszuhalten.