

THE PRIORITY OF TEMPORAL ASPECTS IN L2-SWEDISH  
PROSODY: STUDIES IN PERCEPTION AND PRODUCTION

Bosse Thorén



# **The priority of temporal aspects in L2-Swedish prosody**

Studies in perception and production

**Bosse Thorén**



Stockholm University

©Bosse Thorén, Stockholm 2008  
©Cover picture, Hans Thorén, Vadstena 2008

ISBN 978-91-7155-610-3

Printed in Sweden by US-AB PrintCenter, Stockholm 2008  
Distributor: Department of Linguistics, Stockholm University

To all my lovely students  
who inspired this work



# Contents

Phonetic transcriptions.....	xi
Introduction .....	12
1.2. Foreign accent.....	17
1.2.2. Kinds and degrees of foreign accent .....	18
1.3. A growing interest in prosody in the teaching of Swedish as an L2.....	23
1.3.1. Experience-based ideas .....	23
1.4. Principles of Basic Prosody as guidelines for a teaching strategy.....	27
1.4.1. The development of the present Basic Prosody (BP) .....	27
1.4.2. The Principles of BP .....	31
1.4.3. The postvocalic consonant .....	35
1.5. Some theoretical and pedagogical considerations .....	40
1.5.1. Priority among prosodic phonological features.....	41
1.5.2. Priority among phonetic correlates .....	46
1.5.3. Pedagogical considerations .....	52
1.6. Methodological aspects of duration measurements.....	59
2. Introduction to empirical studies .....	62
2.1. Segmentation .....	62
2.2 Studies aiming at the further illumination of the major issues in the concept of Basic Prosody .....	63
2.2.1. Experiment 1 .....	63
2.2.2. Experiment 2 .....	64
2.2.3. Production study 3 .....	64
2.2.4. Production study 4 .....	64
2.2.5. Production study 5 .....	65
3. Empirical studies .....	66
3.1 Duration versus spectrum: A perception test with manipulated durations.....	66
3.1.1. Introduction .....	66
3.1.2. Method .....	68
3.1.3. Result .....	69
3.1.4. Discussion.....	74
3.2. The postvocalic consonant as a complementary cue to the perception of quantity in Swedish – a revisit.....	76
3.2.1 Introduction .....	76

3.2.2 Method .....	79
3.2.3. Result .....	81
3.2.4 Conclusion and discussion .....	82
3.3. L2-speakers' strategies to convey the Quantity contrast in Swedish – a production study .....	84
3.3.1. Introduction .....	84
3.3.2 Method .....	89
3.3.3. Results .....	91
3.3.4. Conclusion and discussion .....	96
3.4. L2-speakers' strategies to convey the word stress contrast in Swedish – a production study .....	100
3.4.1. Introduction .....	100
3.4.2. Method .....	103
3.4.3. Results .....	105
3.4.4. Discussion.....	109
3.5. Swedish Accent – Duration of postvocalic consonants in native Swedes speaking English and German.....	112
3.5.1. Introduction .....	112
3.5.2. Method .....	113
3.5.3. Results .....	114
3.5.4. Discussion.....	116
4. General discussion and preliminary conclusions .....	119
4.1. Priority of certain aspects of L2-phonetics over others .....	119
4.2. Priority of temporal aspects over tonal.....	122
4.3. Priority of temporal aspects over spectral.....	123
4.4. The inclusion of postvocalic consonant “length” in a pedagogic approach .....	126
4.5. The suitability of BP as basic principle of pronunciation teaching for L2-learners of Swedish .....	129
Summary in Swedish (sammanfattning på svenska).....	132
Inledning.....	132
Bakgrund.....	132
Principer för basprosodi .....	134
Prioritering av temporala fonologiska kontraster framför tonala.....	135
Prioritering av temporal realisering av betoning och kvantitet framför tonal respektive spektral .....	136
Empiriska studier.....	137
Duration eller spektrum som korrelat till kvantitet.....	138
Den postvokaliska konsonantens duration som en kompletterande perceptuell ledtråd till kvantitetsdistinktionen .....	139
Andraspråkstatares strategier för realisering av svenskans kvantitetsdistinktion .	139
Andraspråkstatares strategier för realisering av svenskans betoningsdistinktion .	140



Svensk brytning – den postvokaliska konsonantens duration hos svenskar som talar engelska och tyska .....	141
Allmänna slutsatser och diskussion .....	142
Acknowledgements .....	143
References .....	144

# Abbreviations

L1	first language
L2	second language
BP	Basic Prosody
FA	Foreign accent
SLM	Speech Learning Model
Sw.	Swedish
V	vowel general sense
V	vowel in unstressed syllable
C	consonant
W	word: in measures of relative durations
U	utterance
M	mean value
Sd	standard deviation
S	strong/stressed syllable
W	weak/unstressed syllable: stress patterns, only in 1.4.2.
ADS	adult directed speech
IDS	infant directed speech
Ms	milliseconds
M	mora
$\Sigma$	syllable

# Phonetic transcriptions

Transcriptions in this thesis will use IPA characters in a semi-phonemic way in those cases where there is not a special interest in a specific allophone. This means that e.g. [r] will be used although many realizations of this phoneme in everyday speech are approximants [ɹ] or fricatives [ʀ], and that aspiration is disregarded. Swedish /a/ and /e/ in unstressed positions are transcribed as respective schwa-character [ə] and [ɐ]. Prosodic marking in polysyllabic non-compounds will indicate word accent category and phonologic length, e.g. *äta* ‘eat’ is transcribed as [ɛːtə], *kommer* ‘comes’ as [kóm:ər]. Monosyllabic words will not be marked for stress allocation. In compounds and other words containing two stressed syllables, main stress is marked with word accent character over the vowel in the main stressed syllable, and secondary stress is marked with standard character before the secondary stressed syllable, e.g. [pötò:tɪs,mu:s]. There is a variety of Swedish /ʉ/, ‘Bergslags-u’, common e.g. in Finland-Swedish, which is in many contexts transcribed as [ʉ]. In the present thesis however, it has to be distinguished from central Standard Swedish [ʉ], and is therefore given the diacritic of ‘mid-centralized’ [ǎ]. In quotations from other authors, and in borrowed figures and tables, the transcriptions are faithful to the original.

# Introduction

"And the Gileadites took the passages of Jordan before the Ephraimites: and it was so, that when those Ephraimites which were escaped said, Let me go over; that the men of Gilead said unto him, Art thou an Ephraimite? If he said, Nay; Then said they unto him, Say now Shibboleth: and he said Sibboleth: for he could not frame to pronounce it right. Then they took him, and slew him at the passages of Jordan: and there fell at that time of the Ephraimites forty and two thousand." (Judges 12:5-6, KJV)

Pronunciation in a second language is not often a question of life or death, as in the passage above, but it may well be of great significance when people use languages other than their native ones. The pronunciation may influence both how well the speaker will be understood, and how he/she will be treated otherwise. Second and subsequent languages have probably been learned and taught as long as there have been speaking people in the world, and the notion of foreign accent is probably as old. Foreign accent is a phenomenon that is ubiquitous, almost everyone who speaks a second language – that was not learned in childhood – has it, and it does not seem to be connected to the knowledge and mastery of vocabulary, morphology or syntax. Statesmen and highly educated individuals who know the language very well still have accents – Joseph Conrad, Henry Kissinger and Theodor Kallifatides (Greek-Swedish author) to mention a few.

Nowadays much of the learning and teaching of non-native languages takes place in institutions, with teachers specialized in second language pedagogy. The teachers may be native speakers of the target language or highly skilled users of the target language as a second language (henceforth L2). The teachers have probably studied vocabulary, grammar and phonetics of the target language and if they have learned it as an L2, they are probably fluent in the language.

Every teacher meeting beginners who have no knowledge of the target language, are confronted with the challenge of choosing the aspects of the language which seem proper to start with and which to emphasize. It may be aspects of the target language that are assumed to be easily understood and easily applied by the learner, and that are assumed to be useful in communication. All areas in language teaching have to go through stages of priority decisions. These decisions can be conscious or unconscious, but they must take place, since “teaching and learning all there is” is simply not possible.

This thesis is a discussion of the groundwork for the formulation of an L2 teaching strategy for Swedish pronunciation, based mainly on simplified descriptions of Swedish prosody and with a focus on temporal properties. It should be pointed out here that the present thesis does not address how to teach, but rather what to teach, i.e. aspects of learning goals. The focus on temporal properties entails in this case also toning down the importance of tonal and spectral properties. The properties are a small selection of all possible phonetic properties, and in the priority of goals that is advocated here, lies the assumption that precisely these chosen properties are crucial for making oneself understood, and also that they constitute a robust basis/foundation to build on, if the learner is aiming at acquiring a “perfect” Swedish pronunciation, i.e. a pronunciation that is easily understood by most categories of native listeners. It should also be stressed here, that earlier suggestions for such groundwork have significantly influenced the discussion in this thesis. Earlier suggestions as well as the approach advocated in this thesis are not the direct result of scientific work. They are however constructed by highly skilled academic phoneticians, who wanted to comply with the increased need for good language teaching methods, as Sweden in the 1970’s received a massive influx of refugees from many parts of the world. A vast majority of the immigrants takes part in free national programs of Swedish instruction. The earlier suggestions for guidelines in the formulation of an L2 teaching strategy for Swedish pronunciation have set the tone for Swedish L2 teaching from the 80’s up until now, and can be considered interesting because of their great influence, and because of the fact that they focus on prosody. The development of the ideas will be briefly described, with numerous references to “experience”, “intuition” and “assumptions”. Despite their lack of scientific stringency, they provide a necessary background to the discussions and empirical studies in the present thesis. Implications and consequences of teaching approaches will be related to existing theories and findings, and eventually some of their components will be tested in empirical studies. The lack of empirical research within the field of L2 pronunciation training does not appear peculiar to Swedish conditions; the corresponding situation for English as L2 is compellingly described by Derwing and Munro (2005).

The aims of simplified pronunciation teaching strategies do not, as indicated above, have to restrict themselves to mere intelligibility, since individual learners are able to reach different levels of mastery, with respect to native-like pronunciation in a second language. The result of L2 pronunciation learning is e.g. strongly dependent on the learner’s age, and a strong correlation between the age of the learner and the phonetic result of the learning process has been shown by Flege et al. (1995), Piske et al. (2001) and Abrahamsson & Hyltenstam (2006). This correlation should however not be regarded as a support for the Critical Period Hypothesis introduced by Lenneberg (1967), since there are still individuals that started their learning as

adults, who acquired native-like L2 pronunciation, which is also in agreement with e.g. Markham (1997), and there are also L2-learners who exhibit a foreign accent in spite of the fact that they started learning the L2 before the age of 6 (Flege et al. 1997, Thompson 1991). Although the number of L2 learners in Abrahamsson & Hyltenstam (2006), who were judged as native speakers of Swedish decreased noticeably between the ages of roughly 12 and 20 years, it is of great significance for the language instruction to adult learners, that there is no absolute critical age over which a person can not learn an accent free L2. The results mentioned opens for the possibility that good instruction can exert influence on the learning outcome. So, provided that we believe that pronunciation teaching will make a difference, there should be guidelines especially in the field of adult L2 instruction, where “accent free” is not a realistic goal for the majority of learners. If the guidelines are optimally defined, they should provide a good basis for intelligibility, and also for further perfection. Optimism within the field of formal L2 instruction is further supported by results from studies in perception training, where adult L2 learners are shown to be capable of rapidly learn to perceive L2 contrasts (Jamieson & Morosan 1986, 1989, MacKain et al. 1981, Morosan Jamieson 1989). Special training in L2 pronunciation has resulted in better pronunciation in the target language compared to general language instruction (Bongaerts et al. 1997, Moyer 1999, and Missaglia 1999).

The present work deals mainly with the rules that describe the language, and the reasons for including a specific sample of phonetic properties in the rules. An important prerequisite for a set of teaching guidelines is that they are robust, and can be applied by teachers and learners who have neither the ambition nor the possibility to become experts in phonetics or phonology. Moreover, a set of educational goals that are not the direct result of empirical studies, must be based on a number of assumptions. It is one of the aims of this thesis to scrutinize some of the assumptions that underlie the proposal of the present thesis, and judge whether they can be scientifically defensible, in the light of existing knowledge and theories about Swedish prosody on the one hand, and theories about L2 phonetics on the other. Another aim is to empirically study variations in the realizations of the Swedish stress and quantity contrasts, which are the main ingredients in the proposed teaching strategy. The package of priorities, goals and descriptions proposed in this thesis will henceforth be referred to as “the Basic Prosody approach”, or “BP-approach” or only “BP”. It is here admitted that the Swedish term *Basprosodi* ‘Basic prosody’ was first<sup>1</sup> used by Bannert (1979b), but unaware of this fact, the present author and a colleague used it in Slagbrand & Thorén (1982, 1997). We can only hope that this theft is outlawed by now.

---

<sup>1</sup> As far as the present author is aware.

The present author entered the field of Swedish as an L2 with a BA<sup>2</sup> in linguistics, and has applied the BP-approach for many years of practical instruction. This thesis proposes and motivates some major aspects of a strategy (BP) for teaching Swedish as an L2. BP focuses on some general issues in L2 Swedish teaching and on the priority of some features of prosody. It is based on intuition and experience as well as modern phonetic research.

One aspect that complicates linguistic description in the pedagogical field, is that comparison to natural L1-speech only, is not satisfying. This in turn depends on the experience (of the present author and many colleagues) that some pronunciation goals require substantial exaggeration from the teacher, in order to elicit an acceptable realization (or imitation) from the learner, than others. This in turn can yield a language description that emphasizes one aspect, e.g. long consonant after short stressed vowel, and tones down another, e.g. lax quality of short vowels in stressed syllables. Not because of unequal importance of the respective goal to render intelligibility to the speech, but because one goal has appeared easier to acquire – or is already present in the learner’s L1 – as an L2-student than the other. This aspect motivates why some tentative suggestions to alternative phonologic analyses are presented in the following sections.

The reasoning and the studies will be centered on Swedish temporal prosody in general, and on the role of the postvocalic consonant duration in stressed syllables in particular. The experiments and production studies reported in the present dissertation will hopefully shed some light on how phonological word stress and quantity, can be realized and perceived in an L2 setting. The studies are also an attempt to test the appropriateness of BP, and of the underlying description and the priorities it reflects. Table 1.1 below gives an overview of Swedish phonology. The present work will deal mostly with columns 1 and 2 in the upper half of the table.

Two main sections can be identified in the phonology of Swedish, and other languages; prosody and segments, which is illustrated in table 1.1, upper and lower part respectively. With respect to prosody, Swedish has three phonologic contrasts; stress, quantity and tonal word accents. Among segments the vowel inventory contains vowels that all contrast with each other, the same goes for consonants. The rightmost lower column does not stand for a category of contrasts, but unites diverse phonological processes, where different segments influence the realization of others.

The two leftmost columns correspond to the two most important pronunciation goals, according to the present author, as well as Bannert (1980), and are also the object of interest in the present thesis. They constitute what in everyday speech is called rhythm, but according to e.g. Fraise (1982: 156):

---

<sup>2</sup> The Swedish exam is “filosofie kandidat” or “candidate of philosophy”, which has BA as its nearest correspondence.

Table 1.1. Outline of main areas and contrasts in Swedish phonology. Dashed frame around the columns most relevant to the present work. Translated from the Swedish version in Thorén (2008).

1. Stress	2. Quantity	3. Tone /Intonation
<p><i>'kanon</i> – <i>ka'non</i> [ˈka:nɔn] – [kə'nu:n] 'canon' – 'cannon'</p> <p><i>'racket</i> – <i>ra'ket</i> [ˈrak:ət] – [rə'ke:t] 'racket' – 'rocket'</p> <p><i>'Japan</i> – <i>ja'pan</i> [ˈja:pən] – [jə'pa:n] 'Japan' – 'japanese'</p> <p><i>armen</i> – <i>armén</i> [ˈar:mən] – [ər'me:n] 'the arm' – 'the army'</p> <p><i>'planet</i> – <i>pla'net</i> [ˈplɑ:nət] – [pləne:t] 'planet' – 'the plane'</p> <p>Stress can be distinctive also on sentence level, e.g:</p> <p><i>'hälsa på någon</i> – 'Greet someone'</p> <p><i>hälsa 'på någon</i> 'Visite someone'</p>	<p><i>glas</i> – <i>glass</i> [gla:s] – [glas:] 'glass' – 'ice cream'</p> <p><i>vīt</i> – <i>vitt</i> [vi:t] – [vit:] 'white' – 'white'</p> <p><i>vila</i> – <i>villa</i> [vi:lə] – [vil:ə] 'rest' – 'villa'</p> <p><i>söt</i> – <i>sött</i> [sø:t] – [søet:] 'sweet' – 'sweet'</p> <p><i>bus</i> – <i>buss</i> [bʉ:s] – [bøs:] 'mischief' – 'bus'</p> <p><i>rot</i> – <i>rott</i> [ru:t] – [røt:] 'root' – 'rown'</p> <p><i>vit</i> – <i>vitt</i> and <i>söt</i> – <i>sött</i> differ in Swedish grammatical gender.</p>	<p>Terminal tone ↙</p> <p>Continuation tone →</p> <p>Accent 1 (acute) <i>änden</i> 'the duck' <i>tömt</i> 'the building site' (Receives its tonal properties from the sentence intonation)</p> <p>Accent 2 (grave) <i>anden</i> 'the spirit' <i>tömt</i> 'Santa Claus' ↘</p>
4. Vowels	5. Consonants	6. Assimilations and the like
<p>"Hard"/Back vowels /a u ʊ* o/</p> <p>"Soft"/Front vowels /e i y ɛ ø/</p> <p>Qualitative/ Spectral difference between long and short allophone</p> <p>Substantial for /a / and / ʊ /</p> <p>Negligible for /ɛ/ and /ø/</p> <p>*/ʊ/ is not a back vowel in a strict sense, but is grouped in the back vowel class because of similar impact on certain preceding consonants.</p>	<p>Voiced /b d g j l m n r v ɳ/</p> <p>Voiceless /f h k p s t ʃ/</p> <p>Phonotax: Swedish allows initial clusters of 3, e. g. <i>skriva</i> 'write' and finale up to 5, e.g. <i>skämskt</i>, 'roguish'</p> <p>but in reality, a maximum of 3-4 finale consonants are realized.</p>	<p>/r/ + dental consonant = supradentals (retroflex sounds) <i>bord</i> [bu:d] <i>svart</i> [sva:t] <i>kurs</i> [kø:s] 'table' 'black' 'course'</p> <p>Voicing assimilation <i>Onsdag</i> &gt; [ˈon:ste]</p> <p>Nasal assimilation <i>En bil</i> &gt; [əm'bi:l]</p> <p>Word boundaries: Liaison <i>Han hete rAnders</i></p> <p>Word-finale r-deletion before consonant: <i>Vad äte(r)Bengt?</i></p> <p>No doubling of identical (or semi-identical) segments: <i>Vad läse(r)Rickard</i> <i>Vet du</i> &gt; <i>Vetu</i> etc.</p>

“All perceived rhythm is the result of an activity by the subject since physically there are only successions”, and “temporal organization”, “temporal pattern” or “temporal structure” are the terms that will be used most often in this work.



To summarize: A strong interest in prosody, in the field of teaching Swedish pronunciation to L2-learners has developed in the last 3 decades. Simultaneously there has been the realization that not all phonetic features are equally important to make the L2-pronunciation intelligible and “listener friendly” in the field of adult instruction, where native-like L2-pronunciation is attainable for very few learners. The proposal of the present thesis is influenced by other strategies that were roughly simultaneously developed by researchers on mainly intuitive grounds. The present strategy (BP) is based on a simplified description of Swedish prosody, with emphasis on temporal properties. The strategy and its underlying description will be introduced and related to theories and findings in Swedish prosody. A discussion on details and consequences derived from the strategy follows and thereafter reports from five studies; two perception experiments and three production studies. Finally, some general conclusions and discussion will follow. It should be pointed out that the present work does not deal with teaching methods or pedagogical theory, only with the appropriateness and usefulness in priority and descriptions, with special interest in temporal correlates of Swedish stress and quantity.

## 1.2. Foreign accent

### 1.2.1. Definition

The notion of foreign accent – that opened this dissertation – can involve grammatical, lexical and phonetic features and gives the listener the impression that the speaker is not native in the language. Sometimes the listener can tell what language the speaker’s L1 is, and sometimes not. In Crystal (2003: 3) “accent” is defined as “...The cumulative auditory effect of those features of pronunciation which identify where a person is from, regionally or socially”. Crystal distinguishes accent from dialect, as the latter is seen as including vocabulary and grammar in addition to pronunciation. The term “foreign accent” is in Crystal listed among “regional accents”, a classification that is disputable. From one point of view, we can agree, since the accent can tell one Swedish listener whether the speaker is from Skåne (southern Sw. province), Dalarna (central Sw. province), Germany or Russia. From another point of view foreign accent is of another kind, since a speaker presenting foreign accent is per definition a speaker of at least two languages, and the speaker of “Dalarna-Swedish” may have that accent as the only variety in his/hers spoken linguistic repertoire. Foreign accent can be regarded as one or more values among other values, placing the idiolect of an individual speaker on an ordinal scale of linguistic proximity, ranging from the idiolect of one person, to a foreign language that is not understood by the listener.

### 1.2.2. Kinds and degrees of foreign accent

Foreign accent can be quantified in a number of ways; how strong is the global foreign accent? How intelligible is it? How pleasant is it? What do you think of the speaker's educational/economic/cultural/intellectual level etc. when listening to him/her? It is shown e.g. by Boyd (2004) that people tend to ascribe competencies other than phonetic to speakers, based on phonetic deviations from the target language. Boyd (2004) correlated ratings of global foreign accent, general language competence, suitability to teach, grammatical correctness and variation in vocabulary for 5 non-native teachers working in Sweden, with measured deviations from a phonetic Swedish pronunciation norm, and found that all mentioned variables were correlated to deviant pronunciation. The perceived degree of grammatical correctness and vocabulary were better correlated to perceived and measured degree of deviant pronunciation than to objective measures of grammar and vocabulary. This implies that better L2-pronunciation can promote a professional career as well as mere intelligibility of the spoken L2. In Boyd (2004) an attempt was made to compare the impact from prosodic and segmental deviations, on the perceived degree of foreign accent. The total score for phonetic deviations correlated well with perceived degree of foreign accent, but when segmental and prosodic deviations were separated, the number of segmental deviations per 100 words correlated better with perceived degree of accent, than did prosodic deviations. Segmental deviations are defined as "vowels", "consonants" and "idiosyncratic segments", but unfortunately the measurements of prosodic deviations are not specified, and the prosodic deviations are merged into one category and individual contributions from e.g. temporal and tonal features can not be evaluated.

Bannert (1990: 61, 64) points out how some aspects of phonetic deviations in foreign accent, affects how the native listener will perceive the speech:

- The *number* of deviations in the speech, i.e. how many types of deviations.
- The *combination* of deviations. In those cases where a specific consonant is incorrect, the vowel quality and quantity in the same words is also often defective.
- The *frequency* of the deviations, i.e. how often they occur in fluent speech. (Translation and adaptation from Swedish by the present author)

Munro & Derwing (1995) introduce a division of "foreign accentedness" into three listener-related components:

- *Intelligibility*: The extent to which a listener actually understands an utterance. Tested by percentage words correct reproduced. Also correct answer to questions, and the time it takes to reproduce or answer.

- *Comprehensibility*: A listener's perception of how difficult it is to understand an utterance.
- *(Global) Accentedness*: A listener's perception of how different a speaker's accent is from that of the L1 community. (Definitions from Munro & Derwing 2005: 385).

### 1.2.3. Foreign accented perception

So far, we have treated the spoken foreign accent, i.e. speech production containing deviations from the norms of the target language. A reasonable assumption is that one important cause of accented speech is the kind of contact the learner first has with the L2. An L2-input that is purely written may also give rise to a spoken L2 in the learner. In that case, the learner's knowledge of the relation between letters and sounds in the L2 is crucial for the outcome. In the case of spoken input, there may still be sources of distortion between the aural input, via perception, category building and speech planning, to automatic speech production, requiring speech motor patterns that are unfamiliar to the learner. The difficulty of perceiving all relevant categories and contrasts of the L2 is introduced in Strange (1995) and discussed in McAllister (1997), and the term "perceptual foreign accent" is used "to designate the global effect of this perceptual difficulty as measured by a speech comprehension test". McAllister (1997) compares this definition to an earlier use of the same term, by Strange (1995), as "the significant difficulty which adults have perceiving most (but not all) phonetic contrasts that are not functional in their native language". Perception of L2-distinctions and categories is assumed to be a prerequisite for correct production of L2-distinctions and categories. This is supported by at least one study (Rochet 1995), concerning the ability to perceive and produce French /y/, by native Portuguese and English adults. There is however also an example of Japanese L2-learners being able to produce the American English /ɹ/-/l/ contrast without being able to perceive it (Yamada & Tokhura 1992).

The mechanisms of the defective perception of new sounds and new contrasts are accounted for in the Speech Learning Model of second language sound acquisition (SLM) by Flege (1995). It contains 4 postulates and 7 hypotheses, and a couple of its more interesting features are the notion of "equivalence classification" in H5, and the second part of H6, concerning representations of categories based on different features. Hypotheses 5-7 claim:

"H5: Category formation for an L2 sound may be blocked by the mechanism of equivalence classification. When this happens, a single phonetic category will be used to process perceptually linked L1 and L2 sounds (diaphones). Eventually, the diaphones will resemble one another in production.

H6: The phonetic category established for L2 sounds by a bilingual may differ from a monolingual's if: 1) the bilingual's category is "deflected" away

from an L1 category to maintain phonetic contrast between categories in a common L1-L2 phonological space; or 2) the bilingual's representation is based on different features, or feature weights, than a monolingual's.

H7: The production of a sound eventually corresponds to the properties represented in its phonetic category representation." (Flege 1995: 239)

The SLM seems optimally adapted to categories and contrasts on phoneme level, but "category building" is relevant also for prosodic contrasts and categories. An L2-learner of Swedish could e.g. perceive a quantity category based on vowel spectrum, vowel duration, vowel over consonant duration or combinations of these. The different alternatives are assumed to result in different representations of categories in the learner, and hence – according to H7 – different productions. The categories created by the Swedish quantity contrast could be assumed to be the subject of equivalence classification if the learner has American English as L1. The Swedish /V:C/-category is similar to the English 'tense' category, and could be expected to be classified as such, and the Swedish /VC:/-category could be expected to be classified as 'lax'. The English tense-lax contrasts rests mainly on spectral correlates (Garlén 1988, McAllister et al. 2002), whereas the Swedish quantity contrast rests mainly on temporal correlates (Behne et al. 1997, section 3.1, this work), including complementary consonant duration. Thus, a prosodic contrast can be expected to be perceived by the American English learner of Swedish, classified as a familiar L1 contrast, and realized according to the L1 pattern. This is in agreement with the experience of the present author, especially in the case of American English learners. In this case, the contrast would be realized mainly with spectral means, and small – if any – temporal means, whereas the weights of the two correlates in Swedish is the opposite, as mentioned above. In the typical case, there would be no increase in post-vocalic consonant duration after short vowel allophone. This would correspond to case No 2 in H6 (Flege 1995) above. An L2 teacher who is aware of this risk may help the learner to focus on the correct features, or the correct feature weights of realization. Sections 3.3 and 3.4 report production studies in which L2-speakers realize Swedish word stress and quantity. These productions are analyzed with respect to acoustic correlates, with special focus on temporal cues. In study 5 (3.5) it is also shown how the Swedish complementary pattern (V:C – VC:) is present when native Swedes speak English and German.

#### 1.2.4. Foreign accent in a pedagogical perspective

The division of foreign accentedness, into intelligibility, comprehensibility and global accentedness, presented in 1.2.2 above, can be seen as highly relevant to the pedagogic field, since this field needs guidelines as to 1) what phonetic features are crucial to master in order to be understood in the particular target language 2) what phonetic features will be productive in taking

the learner not only to “intelligibility-level”, but also to listener friendliness. We can state that there is not much research done in this field, so we are still dealing with intuitive conclusions and assumptions although the intuition and the assumptions are tempered by long experience and specialized education in linguistics and phonetics. The present author has communicated with native Swedes who react negatively to regional varieties of native Swedish spoken in radio or television, and who also react negatively to phonetic details revealing that speakers in radio or TV have other L1’s than Swedish. He has however also heard a large number of teaching colleagues agree that certain prosodic elements in L2 Swedish tend to conceal many segmental deviations. It seems logical to assume that mere intelligibility is not enough to promote integration, since subjective emotional judgments will be made by the listener. One assumption underlying the BP-approach is that elements of L1-phonology in the L2-Swedish are totally harmless per se, as long as certain aspects<sup>3</sup> of the pronunciation are realized according to the rules of the target language. This means that one variety of typical Polish accent in spoken Swedish – e.g. one that realizes stress and quantity in “the Swedish way” – can be quite acceptable, while another variety of typical Polish accent in spoken Swedish – e.g. one that strictly follows the Polish stress pattern, and does not realize the Swedish quantity distinction at all – can be almost unintelligible. Both kinds of Polish accent could have the same segmental deficiencies. According to Munro & Derwing (1999) the *presence* of foreign accent decreases intelligibility, but *degree* of global foreign accent is not correlated to degree of intelligibility.

An adult learner of Swedish as an L2 is likely to speak with a foreign accent also after considerable length of residence in the L2 environment (e.g. McAllister 2001), and this accent may be acceptable to some listeners and annoying to others, since emotional attitudes towards different L1’s and towards different phonetic deviations obviously vary among both native and non-native Swedish listeners. The teacher has no guarantee that the chosen pronunciation goals will promote general listener friendliness to a population of listeners, but there is – according to the experience of the present author – a growing consensus that prosody is a crucial phonetic and phonological property, that helps in mitigating the negative reaction to any perceived traces of certain L1’s.

It is a strong impression of the present author, that in the Swedish debate – among non-phoneticians – on how foreign accent results in worse integration, poor job careers etc. there are three main issues being discussed: 1) how strong is a person’s foreign accent? 2) what is the status of the L1 reflected by the foreign accent? 3) The Swedish listeners must become more tolerant and more willing to understand foreign accents. This could be interpreted as 1) the less foreign accent, the better, and 2) a speaker with an e.g. typical

---

<sup>3</sup> It is not possible to exactly define these aspects, but a thorough discussion of this and related issues, is presented in section 1.5.

American English accent will be treated better in Sweden, than a person with an e.g. typical Arabic accent. 3) Less efforts to develop good pronunciation teaching methods. As mentioned above, one of the points of the present thesis is that certain phonetic details in any foreign accent will determine how well the speaker will be understood and otherwise perceived. As is also mentioned, two speakers presenting very typical e.g. Polish accent, can still differ very much in how they master more or less crucial phonetic features in Swedish. This reflects the idea that specific kinds of realizations of specific phonological features are more determining for a foreign accent's intelligibility and listener friendliness, than degree of global accent, or the foreign accent typically reflecting a specific L1. We should consequently stipulate a dividing line between *naturalness*, meaning closeness to native-like speech, and *listener friendliness*, meaning that traces of the L1 can be present, but with high degree of intelligibility and low degree of features that tend to disturb the Swedish listener, or distract the listener's attention from the content of the spoken message. This distinction between *naturalness* understood as degree of perceived similarity with native speech, and *listener friendliness* is an attempt to defend the priority of certain phonetic properties at the cost of others. The division itself does not say anything about the validity of the priority. Figure 1.1 below is an attempt to illustrate how the tentative construct of listener friendliness is more depending on intelligibility than on degree of global accentedness. The latter variable is assumed to be equal to naturalness, which in turn is interpreted as degree of native-likeness. "Listener friendly" could hypothetically cover the whole range of naturalness,

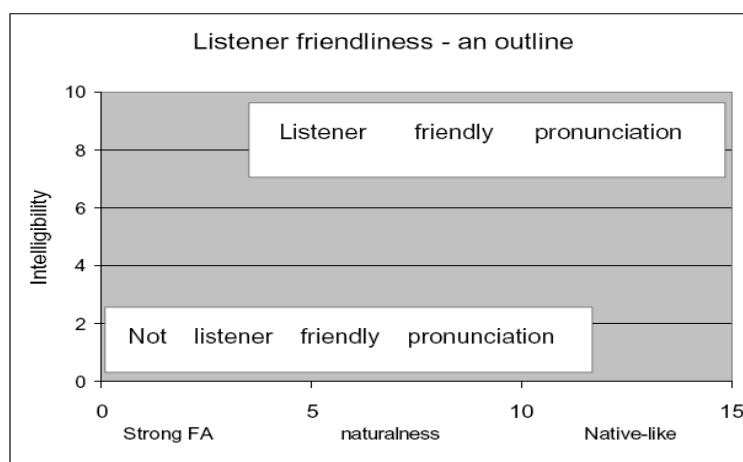


Figure 1.1. Listener friendliness in relation to intelligibility and naturalness. Scales on x and y-axes are arbitrary but ordinal.

but the uncovered range of "strong global foreign accent" is assumed as unlikely to be perceived as listener friendly. There is a similar assumption in the case of "not listener friendly", which could go all the way to "native-

like”, since a person can speak with a native but odd regional accent, having something in the mouth, speaking thickly as a result of illness or intoxication or all at once. The figure emphasizes once more that eliminating all traces of the L1 is probably not the optimal pronunciation goal for the adult L2-learner. Whether this assumption is correct or not, is another issue, but it is one of the ideas underlying the origin of the BP-approach and other simplified teaching approaches.

### 1.3. A growing interest in prosody in the teaching of Swedish as an L2

#### 1.3.1. Experience-based ideas

One can hardly claim that the learning of temporal prosody is more important for intelligibility or “listener friendliness” than a sequence of recognizable phonemes. In the educational approach to be examined lies however an experience-based assumption that all or most L2 learners spontaneously strive to produce a “correct sequence of segments” rather than giving the sequence “L2 prosodic properties”.

In Malmberg (1967) there is a passage that does not claim prosody to be of higher interest than e.g. segmental phonemic contrasts, but stresses that prosody must neither be forgotten in L2-teaching, nor postponed:

“One must not deliberately, let the students during the first time of the course, pronounce new sounds with L1-prosody, believing that accents can be acquired at a later stage. This is worse than throwing the yeast into the oven after the dough. It is, like when somebody was to dock the tails of the puppies, cut them piece by piece, thinking this would hurt less.” (Malmberg 1967, 109-110, translation by the present author)

Kjellin (1978) has a paragraph about linguistic micro and macro (see table 1.2). Kjellin compares prosody to the carrier frequency in radio broadcasting, and is of the opinion that prosody is the carrier frequency of the spoken message. He also claims that it is hard to identify the message if you have to tune in to an untrained or unfamiliar frequency. Hence Kjellin claims that the speaker has to adjust the carrier frequency in order to make him understood in the speaking community. In the same chapter Kjellin (1978: 16) writes:

Make sure that you first of all speak the new language with a correct prosody as a basis, and that the syntax is correct. Then gradually practice the correct vowels and consonants, if they have not already become correct from the correct prosody. Eventually the morphology can get some attention (Kjellin 1978, 1992 edition: 16, translation by the present author)

Table 1.2: Macro and micro levels of pronunciation and grammar respectively. From Kjellin 1978: 12 with permission (translation by the present author)

	MACRO	Micro
Pronunciation	PROSODY	Segments
Grammar	SYNTAX	Morphology

This agrees well with Bannert (1979c: 132): “By structuring the segments into larger units, namely syllables, phrases and sentences, the prosody facilitates for the listener to process the flow of sounds, and interpret it linguistically”. The common denominator of both these descriptions seems to be that prosody is a suprasegmental structure onto which smaller units (as segments) are hung, and they are mutually dependent – without segments, nothing to give structure, and without prosody, no structure.

Gårding (1979:13) describes what is assumed to happen when the native Swedish listener is confronted with a word that has a distorted stress:

How about [cəlá:re] (in stead of [cəl:ərə] ‘cellar’)? The word loses its identity. The listener searches for a similar word, i.e. a word with the same stress pattern, rummages around in the brain-lexicon, but finds no correspondence/equivalence. As you see (Gårding refers to a table) it is supposed to mean källare ‘cellar’. (Translation by the present author)

Lindblom et al. (1981: 3) adds fuel to the motivation of investigating the temporal organization of Swedish for pedagogical aims:

Further indication that sound patterns cannot be satisfactorily described if timing is neglected comes from cross-lingual comparisons. In languages such as Swedish and English certain words may have identical stress pattern descriptions but may nevertheless exhibit clearly audible timing differences. Abercrombie's (1964) analysis of English syllable quantity provides us with several examples that can be contrasted with related words in Swedish. Listen to native speakers of English and Swedish pronouncing *shilling* in their respective languages. These disyllables have a main stress on the first syllable but still sound different. The differences have to do with segmental shape and with timing. Compare the Englishman's pronunciation, which we can transcribe as [ˈʃɪlɪŋ] with the Swedish [ˈʃil:ɪŋ]. Typical mistakes found in the pronunciation of English by Swedes are [ˈæt:əm ˈprɪt:ɪ ˈbæl:kɒnɪ] for 'atom', 'pretty', 'balcony'. Although the Swedish speaker may learn to master [ʃ] and [ɫ] in *shilling*, she often tends to make the first vowel and the dark [ɫ] long in accordance with the quantity patterns of Swedish stressed syllables (Elert 1970). No doubt the temporal deviation contributes to giving the English listener an impression of a Swedish accent. These illustrations motivate the conclusion that timing can be an important dimension in contrastive analysis and second language learning.



The priorities of the pedagogical approach to be scrutinized are influenced by a research project in Lund (Bannert 1979b, 1980, Gårding & Bannert 1979), and an instructive book by an enthusiastic teacher and researcher in Uppsala (Kjellin 1978), and the present author's own knowledge and experience of teaching Swedish as a L2 (nearly 30 years). Many teachers of Swedish as an L2, including the present author, intuitively felt that the new way of laying focus on prosody gave better results than before, and made the students sound much more "Swedish" also when producing utterances of some length. Central to this thesis is the priority of consonant length (not made by Bannert), and toning down the importance of the distribution and realization of tone, a domain that is thoroughly elaborated by Kjellin.

As pointed out earlier, the ideas presented here, though partly arisen within a research project, are based mainly on intuition and teaching experience. There are however some more recent studies which support the interest for prosody within the field of Germanic languages as L2's.

### 1.3.2. Empirical support for prosody oriented teaching

First it should be noted that different target languages may profit from different priorities with respect to teaching strategies. Assuming, following e.g. Lindblom (1981), that timing is an important dimension in the case of Swedish as an L2, we cannot automatically assume that the same is true for e.g. Chinese or Polish as target languages. Since Chinese is a tone language with no phonologic quantity, and Polish also lacks phonologic quantity, there may be quite different features that should be emphasized in the pronunciation teaching. There will however be careful parallels drawn between English, German and Swedish, since they are Germanic languages that can have lexical stress on different syllables counted from the beginning or end of the word, and they all exhibit durational differences in a vowel quantity (Swedish, German) or a Tense-Lax (English) contrast. The latter is sometimes described as a quantity contrast, and all three languages realize the phonological contrast with both temporal and spectral correlates.

Bannert (1986) found that the phrase *I samhället* 'In the society' when produced by an L2-speaker, with main stress on *-häl-* only, instead of the correct main stress on *-sam-* and secondary stress on *-häl-* was perceived by native Swedish listeners as other phrases analogue to the stress pattern in the L2 production. When the stress pattern was digitally corrected, the native Swedish listeners tended to perceive the phrase correctly more often. There are also some well known Swedish rhymes<sup>4</sup> – very hard to translate – where the word stress is changed in every word, a distortion that makes the jingle totally unintelligible to the native but naïve listener.

<sup>4</sup> *Mu'lett, ku'lett I'la vu'lett I'de dö'das vi'lorum* sounds like something in Latin to most Swedish listeners. It is however the phrase *'Mulet 'kulet 'illa 'vulet 'I de 'dödas 'vilorum*, meaning roughly 'Clouded, raw, badly ..... in the tomb (rest room of the dead?)

Field (2005) reports decreased intelligibility when lexical stress was shifted in a number of English words that were transcribed by both native English listeners and non-native users of English. The distorted lexical stress caused lower intelligibility for both groups.

Tajima et al. (1997) used short English phrases spoken by Chinese speakers and native English speakers. The L2 productions were temporally corrected by means of LPC resynthesis and dynamic time warping to fit the temporal pattern of the native English production, and the native English productions were correspondingly temporally changed to fit the foreign accented productions. Intelligibility increased significantly in the former case, and decreased significantly in the latter.

Hahn (2004) let native English students listen to three recorded oral presentations with foreign accented English. One version had correct “primary stress” in phrases, one had misallocated stress, and the third had no clear stress in the phrase, i.e. did not give distinct prominence to the most important word in the phrase. The study obviously deals with allocation of sentence stress or focal accent, and the concept of lexical stress is not discussed in the article, leaving the reader unsure with respect to how non-focal lexical stress was realized. Students were tested with respect to recollection of the content, and reaction time in a secondary task, in this case to click a computer mouse whenever a tone was randomly presented during the oral presentation. Her results indicated that correct “primary stress” yielded better intelligibility rates than no primary stress, which in turn was more intelligible than incorrect primary stress. The results were however only partially significant.

Moyer (1999) found that native speakers of English, learning German as an L2, were rated as more native sounding, when they received both segmental and suprasegmental training. The variable “segmental and suprasegmental feedback” was identified as one of the influential parameters with respect to degree of foreign accent in the subjects.

Missaglia (1999) compared two groups of native Italian learners of German. A group that had received “prosody-centered” training improved their pronunciation significantly more than a group that had received “segment-centered” training.

Munro (1995) presented low-pass filtered speech to naïve listeners with English as L1. The speech was produced by native English speakers and Mandarin-speaking learners of English. The filtering had rendered the speech samples unintelligible. The mandarin-accented speech yielded consistently higher ratings for foreign accentedness than did the productions of native English speakers. “...the results suggest that untrained listeners are able to identify foreign-accented speech on the basis of nonsegmental information alone...” (Munro 1995:17).

As stated by Piske et al. (2001), the research on L2 pronunciation learning has been mostly about segments, i.e. vowels and consonants, and contrasts

between segments. There will however appear attempts in this work, to apply findings and models that address category building, to prosodic categories.

To summarize: One research project “Optimization of Swedish pronunciation” resulted – based on researchers’ judgment rather than empirical evidence – in the recommendation of prosody – especially stress and quantity – as a central tool for the teaching of Swedish as an L2, and roughly simultaneously, a book was issued: *Svensk prosodi i praktiken* ‘Swedish prosody in practice’ (Kjellin 1978). Both the project and the book had involved a “pronunciation clinic” activity, and both persons had recently written a doctoral thesis containing substantial amounts of prosody (Bannert 1976 and Kjellin 1976, 1977). Both saw prosody as a most useful tool for the pronunciation teaching in Swedish as an L2. The present work will argue for a priority of goals for Swedish pronunciation, that is influenced by main ideas from Bannert (1979, 1980) and Kjellin (1978), but this thesis emphasizes the temporal aspects of the prosodic system of Swedish. In addition to the experience-based ideas of Bannert, Kjellin and the present author, a number of studies were presented, which gave some empirical support for the idea that misallocated and unclear stress could cause decreased intelligibility, and that prosody centered pronunciation training gave significantly better results with respect to phonetic learning outcome, than segment-centered training. It is not said above that prosody would be of greater significance to Swedish as L2 pronunciation learning than segmental properties, but what is presented may be a reaction to a long period without any structured description or strategy for the use of prosody as a teaching tool.

## 1.4. Principles of Basic Prosody as guidelines for a teaching strategy.

### 1.4.1. The development of the present Basic Prosody (BP)

The idea that all phonetic features do not equally contribute to a good L2-pronunciation, requires a definition of “good L2-pronunciation”. Native-like pronunciation is of course the ultimate “good L2-pronunciation”, and “intelligible” is a reasonable minimally acceptable level. But what if the intelligible pronunciation contains phonetic elements that disturbs or offends the native listener? If nobody manages to listen to the intelligible foreign accent, the message will all the same not get through. A quotation from Abercrombie (1949:120) can illustrate this problem: “Language learners need no more than a comfortably intelligible pronunciation”. This can be interpreted as intelligible to the listener, and comfortable to the listener. We can only hope that the pronunciation is comfortable to the speaker in addition. Let us call

this “comfortably intelligible pronunciation” a listener friendly pronunciation, and also claim that the notion of “comfort” in the intelligible pronunciation is not trivial. How much effort should the listener put into understanding the spoken message? As indicated in 1.2.4, the degree of global accent is not considered a relevant variable with respect to listener friendliness. The present BP-approach claims that correct temporal realization of stress and quantity in Swedish is a prerequisite of listener friendliness, i.e. a comfortably intelligible Swedish.

Do we know what the minimal set of phonetic properties is, in order to speak intelligible Swedish? The answer is no. Sounds can be omitted or distorted. Distorted in how many ways? Contrasts can be neutralized or realized by means of other phonetic correlates than by native users. Do we know what takes an L2-speaker from “uncomfortable intelligibility” to listener friendliness? The answer is no, in spite of many experience-based opinions. This means we still have to be guided to a great extent by intuition, but we can obtain as much empirical evidence as possible in this field, and give more substance to existing pedagogical approaches. We can however state that a listener friendly L2 pronunciation implies intelligibility, and that an intelligible L2 pronunciation may be, but is not necessarily listener friendly. The reasoning can result in a tentative implicational hierarchy.

Native-like >> listener friendly >> intelligible

Since the implicational hierarchy is meant to be a one-way implication, “intelligible” is not necessarily “listener friendly”, and “listener friendly” is not necessarily “native-like”. This construct is meant to justify an intermediate level of expected phonetic learning outcome, in those cases where a native-like L2-pronunciation is not likely to be the result e.g. in the instruction of adults (cf. 1.1.). To illustrate the development of the priority ideas further, two quotations are presented, one from Bannert (1984) and one from Gårding (1979):

Many attempts have been made to improve pronunciation when learning a foreign language, and in these attempts linguistic correctness has been the guiding principle. It seems, however, that hardly any consideration has been given to the native listener’s problem of understanding foreign accent. (Bannert 1984:7)

The pronunciation teaching should not restrict itself to corrections of isolated errors, many of which are of minor significance for the intelligibility, e.g. [mɔ̃:ga] ‘many’ instead of [mɔ̃:a]. An optimal instruction should aim at providing the learner with a good prosodic basis that can support the syllabic and segmental structure. With such a basis, it is possible to speak smoothly and be easily understood, in spite of occasional phonetic errors. (Gårding 1979:18, translation by the present author)

Unfortunately, there are no references in Gårding's text, and it seems that the evaluation of how harmful or harmless different phonetic errors are, is made by two researchers, one of whom has Swedish as an L2. These mostly intuitive judgments seem typical for the way the prosody-oriented teaching ideas have developed, (recall Munro & Derwing 2005, from 1.1)

The notion of "linguistic correctness" above (Bannert 1984) can be interpreted as the ambition that everything in the L2-learner's speech should be identical to the L1 norm. This in turn, means that all phonemic categorization in the L2-learner must be identical to that of the L1-user, with same phonetic realizations of all phonological categories and contrasts. As stated in 1.1 the totally native-like pronunciation is not a realistic goal in the case of most adult L2 learners, and the teacher should not – and probably does not want to – pick out a few features at random, and say, "This is what you should learn to pronounce, since we don't have time to cover more". A selection of phonetic properties will be made, either from the teacher's (or the learner's) intuition or experience, or from scientific work that the teacher (or the learner) has profited by in their education, vocational or other. The more scientific support there would be in such a selection, the better. Figure 1.1 below is an attempt to illustrate the relation between naturalness (or degree of global foreign accent), intelligibility and the tentative construct "listener friendliness".

This idea of communicative efficiency is continuously replacing an older pedagogical way of thinking. This is a strong impression of the present author after contacts with hundreds of teachers of Swedish as an L2, in Sweden and other countries, for more than a quarter of a century, and after reading most manuals on Swedish as an L2. In this older approach seems to lie the assumption, that any deviation from native-like speech must be cured. Recall the discussion in 1.2.4 about naturalness and listener friendliness. Although clear evidence of these two perspectives on L2-pronunciation cannot be provided here, a comparison will probably illustrate two ways of thinking, with bearing on the economics of L2 education. According to a view that accepts only native-like pronunciation, it would be quite logical to put much effort in the curing of all features, and probably most effort in the "hard-learned" features. Among the latter are typically found those phonetic features that are more or less unique to Swedish. And consequently, the more difficult they are to acquire, the more effort must be put into the teaching of them. What if the acquisition of one of these "hard-learned" features does not contribute more than marginally to intelligibility or listener friendliness? And what if a number of easily learned features are neglected in spite of the possibility that they may play a major role for intelligibility and listener friendliness? It could hypothetically be the case that a "non-unique" feature, if acquired, would contribute much more to the L2 speaker's listener friendliness. Below are given three examples of traditionally "hard learned and uniquely Swedish" features, which have no prominent place within BP.

- The tonal word accents 1 (acute) and 2 (grave). See table 1.1, column 3 for an example. A lexical tone contrast that is morphologically predictable, and is distributed and realized differently in different dialects; in singing and some dialects neutralized. More about them in 1.5.1. And 1.5.2.
- A set of three sibilants (in addition to /s/), with phonemic contrast between /ɕ/ and /ʃ/ and semi-free allophonic variation between [ʃ] and [ɕ], although [ɕ] and [ʃ] are acoustically more similar than [ʃ] and [ɕ]. They are listed among less important pronunciation goals in Bannert (1980), and according to the present author's experience, confusion of /ɕ/ and /ʃ/ do not cause any communicative problems as long as /ʃ/ is realized as [ɕ] or [ʃ].
- A greater number of vowel phonemes than most immigrant languages (e.g. Garlén 1988), and in the vowel inventory three front rounded vowels, a category that is relatively unusual in the languages of the world (Ladefoged and Maddieson 1996: 292-297). Swedish children often use a "secret language" replacing all vowels by e.g. /i/ and still understand each other. This indicates that the spoken communication can tolerate some reductions of the vowel system.

So, if there are two philosophies, one that relates L2 pronunciation only to "native-likeness", and another taking listener friendliness into consideration (cf. 1.2.4), the latter would appear as the most attractive.

Before continuing, it should be noted that the term "length" is in the following pedagogically oriented sections intended to refer to both phonological length and physical duration. In the pedagogical setting, this is assumed to be an advantage, since neither teachers nor learners can be expected to always keep phonological distinctions and their phonetic correlates apart. Length and duration are concepts assumed to be intuitively very close to one another, and the phonetic goal of the description is increased duration. When proper, and for the sake of clearness, the terms "phonological length" and "duration" will be used.

The main principles of the preceding Bannert and Kjellin approaches are compared to the present proposal in table 1.3 below. It points out the parts that distinguish the approaches, namely rules for tonal organization, the postvocalic consonant, and vowel quality.

Table 1.3. Points of agreement and disagreement in the three presented strategies

	Bannert	Kjellin	BP
Word stress is high priority	Yes	Yes	Yes
Stress should be realized as increased syllable length	Yes	Yes	Yes
Postvocalic consonant length and vowel length are equally important	No	Yes	Yes
Correct vowel quality to quantity category	Yes	Not included <sup>5</sup> , not denied	Not included, not denied
Tonal organization on word and phrase level	Yes, but not specified	Yes, detailed	Not included, not denied

### 1.4.2. The Principles of BP

The present thesis argues for a simplified pedagogical description of Swedish prosody, that emphasizes stress and length. A central ingredient in the description is the complementary length relation in stressed syllables in Swedish /V:C/ - /VC:/. This complementary relation is interesting also in the sense that it contributes to the realization of both stress and quantity, an issue that will be thoroughly discussed in 1.5. BP is assumed to account for the basic temporal properties of Swedish prosody, and has to be combined with other phonetic features to make up a rich repertoire of L2 pronunciation teaching of Swedish. BP can be summarized as three directives:

1. Stress the correct/suitable word(s) in the sentence.
2. Stress the correct syllable(s) in the stressed word(s) by making these syllables longer than surrounding syllables.
3. Lengthen the correct segment (vowel or immediately following consonant) in the stressed syllable.

As can be seen in table 1.3, BP entails no ingredient that is not already present in either the Bannert or the Kjellin approaches. It merely combines priorities from both, mainly by retaining the consonant length (in agreement with Kjellin) and accepting many different kinds of tonal gestures in the realization of word accents (in agreement with Bannert).

One crucial part of the BP is that duration is assumed to be the main phonetic correlate to both stress and quantity, and that pronouncing the phonol-

<sup>5</sup> The many possible vowel qualities of Swedish are mentioned in the preface of Kjellin (1992), but are not included in the instructions directed to the learner.

ogically long segments with longer duration, will, in addition to signaling the intended quantity category, result in increased duration of the “correct” syllables, which in turn would result in signaling stress in these syllables. Stressed syllables should only be found in stressed words, and this means in turn, that word stress and quantity are only potential properties of lexical items. This would make it tempting to reduce the BP directives to only one:

- Lengthen the correct segment.

This minimalistic instruction makes heavy demands on the learner, with respect to knowledge of sentence stress, word stress and quantity categories, but it captures a central idea in the present BP approach, that if the speaker actually pronounces Swedish with extra length, realized as extra duration, on proper segments with respect to the prominence levels mentioned, the listener is assumed to perceive stress on the proper words, stress on the proper syllables in the proper words, and also the intended quantity category. We will however consider the three original BP directives:

1) agrees with both Bannert and Kjellin, and can also be related to the works of Bruce (1977: 9-24) that strictly speaking is more about tone than temporal organization, but it defines four levels of phonologic word prominence within a sentence: unstressed, secondary stress, accentuation and sentence stress (focal accent). All but the first category can be said to carry stress. All of the stressed categories entail one or more formalized gestures with respect to timing or tone. In the example below all four levels are present.

*Kalle gillar att äta potatis*                      'Kalle (pr. name) likes to eat potatoes'

A neutral way of pronouncing the Swedish sentence above would be to assign word accent to *Kalle* and *gillar*, secondary stress to *äta* (Anward & Linell 1976)<sup>6</sup> and focus accent to *potatis*. The conjunction *att* ‘to’ is unstressed. According to Bruce (1977) the quantity category is realized in all the three degrees that carry stress, the word accent category in the two stronger degrees, and a tonal rise in focal accent only (see table 1.3, section 1.5.1).

2) When a non-compound word is realized in one of the three stressed categories mentioned above, one syllable receives main stress, e.g. the first syllable in

***Kalle***                      [kàl:ə]

---

<sup>6</sup> It can be argued that *äta* in this context, in casual speech, could also be realized without stress, since informal tests by the present author show that recorded /V:C/-words and /VC:/-words can be interchangeable in similar positions, without native listeners notice the exchange, e.g. *känna* ‘after’ ‘feel’ *tjäna* ‘pengar’ ‘earn money’.



<i>gillar</i>	[j̥il:ər]
<i>äta</i>	[ɛ̃:tɐ]

The second syllable in *po'tatis* [pötó:tis] 'potato'

For the sake of simplicity, the phonologically long consonants of *Kalle* and *gillar* are regarded as belonging entirely to the first syllable, and this issue will be discussed in 1.5.3 and 3.4.

Most compounds contain a syllable carrying main stress, that precedes a syllable carrying secondary stress, and thus two syllables in one word may carry stress, e.g.

*Po'tatis,mos* [pötò:tis,mu:s] 'mashed potatoes'

The stressed syllable is pronounced with longer duration than surrounding syllables without stress. Thus, for *Kalle*, *gillar*, *äta* the first syllable should have longer duration than the second one, and for *potatis* the second syllable should have longer duration than the first and the third syllable. In this section we only state that the difference in duration should be detectable to a native listener. Measures of segments and discussion about syllable boundaries will be accounted for in 1.6 and in section 2.1.

3) means that in *Kalle* and *gillar*, the /l/ carries the extra duration, as indicated by the transcriptions above. In *äta*, the /ɛ̃/ receives the extra duration, and in *potatis*, the vowel of the second syllable, /a/ is the long segment. Following the standards for qualities of long and short Swedish vowels, the /a/ in *potatis* is transcribed as [a] or [ɒ], but spectral properties of long and short vowel allophones do not play a principal part in BP, and this will be discussed in 1.5 and be addressed in a perception study, section 3.1.

The 3 BP-rules can be translated into a flow chart (figure 1.2 below) illustrating decisions for the assignment of phonological prosodic features that implies some aspects of lexical retrieval, relevant to the temporal organization of Swedish, from sentence level, via word and syllables, to segment level. The flow chart is a simplified outline that only takes prosodic features into account.

To illustrate the variation of temporal prosody on lexical level, figure 1.3 shows the possible output from the hypothetical process in figure 1.2, if a word consisting of 1-3 syllables is the input to the sequence of decisions, and is realized with minimum secondary stress. The four variables of 1) number of syllables, 2) stress allocation 3) simplex or compound, 4) /V:C/ or /VC:/ yield these 28 temporal structures. The figure does not have room enough for translations, but all the words in figure 1.3 are quite ordinary Swedish words, and each word could be replaced by several other ordinary words. The more subtle distinctions within the stressed categories of compounds are shown by means of the IPA-characters for primary and secondary stress. It can be mentioned that adding the parameter of word accent 1 or 2, to figure 1.4, would result in an additional six possibilities. This variation would be possible in

the S W sequences of trochees (S W), dactyls (S W W) and amphibrachs (W S W), in non-compound words. Swedish compounds are always of accent 2 category.

## Decision phases of length assignment

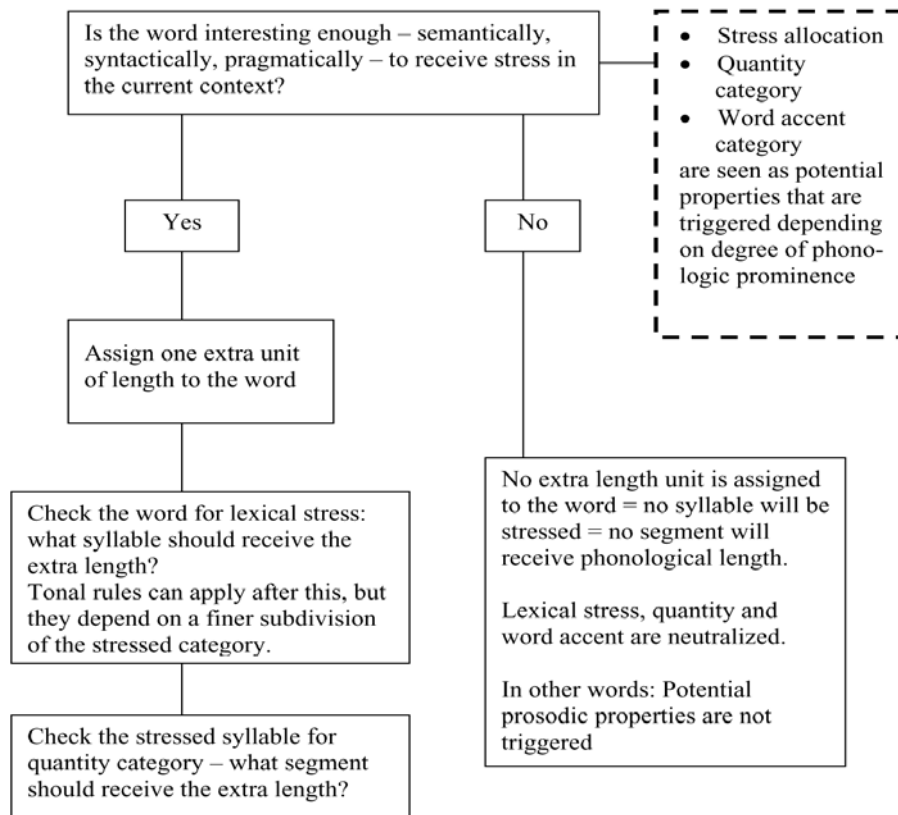


Figure 1.2. Flowchart illustrating decision phases of the assignment of phonologic length, according to the BP rules.

## Possible combinations of length assignment for Swedish mono- di- and trisyllabic words

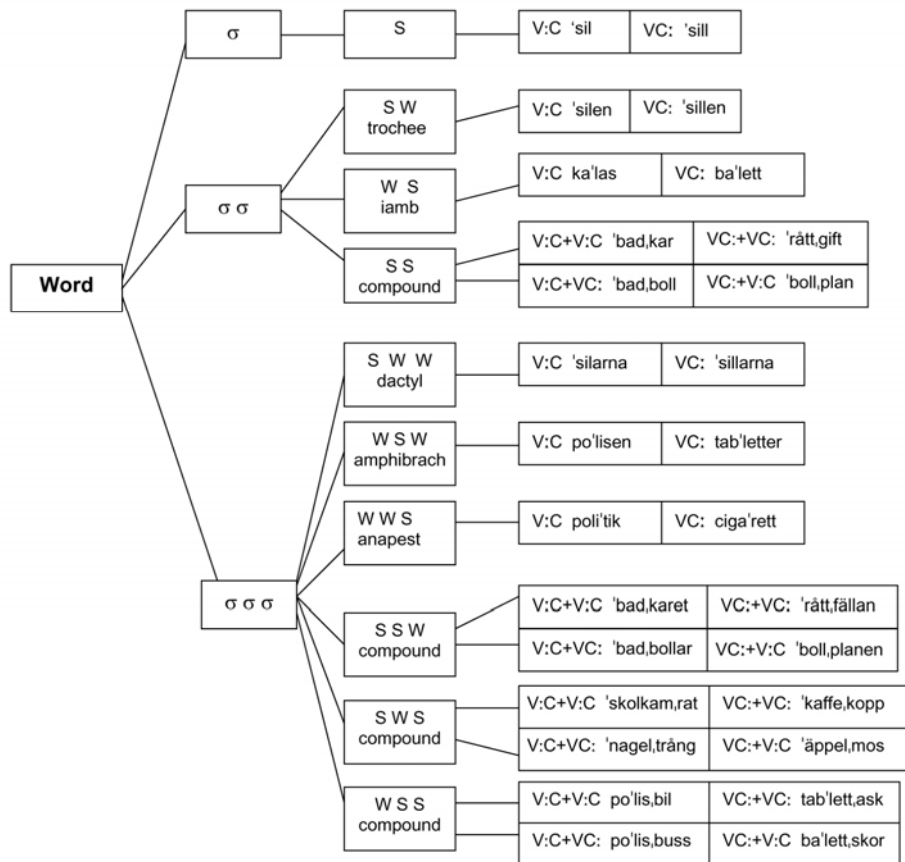


Figure 1.3 Possibilities of stress and quantity structure for Swedish words containing 1-3 syllables.  $\sigma$  = syllable, S = strong (stressed) syllable, W = weak (unstressed) syllable.

### 1.4.3. The postvocalic consonant

The duration as well as the phonologic length of postvocalic consonants in Swedish are necessary logical ingredients in the BP-approach, in order to assign length to a stressed syllable containing a short vowel, but the pedagogical value of this temporal property has been subject to debate within the field of Swedish as an L2. Håkansson & Stenquist (1989:16) declare in their manual on Swedish pronunciation for teachers of Swedish as an L2, in the section that addresses the complementary length pattern: “The length of the

vowel is the most important. The *consonant length follows automatically*. We will henceforth mark only the vowel in stressed syllables.” (Italics by the present author). This view could perhaps serve well in computer programming, since computers do not forget rules you write for them, but we must remember that teaching and learning deal with human beings who cannot be programmed this way. The declaration above could reflect the view that vowel length is something you should learn and that consonant length is less important. It could, however also mean that all learners of Swedish as an L2, have the system of complementary vowel- and consonant length built in as pattern in their speech production mechanism. One must bear in mind that many, perhaps most, languages do not have a duration feature, e.g. a quantity distinction at all, and most languages that do, have either vowel quantity, consonant quantity or both independent of each other. The reader is referred to Schaeffler (2005: 50-53) for a recent survey of quantity in the world’s languages as well as in Swedish dialects. Swedish, Norwegian and Icelandic are known to have vowel quantity with complementary consonant length (Garlén 1988), and Italian is known to have consonant quantity with complementary vowel length (Garlén 1988), both types resulting in the length pattern alternatives /V:C/ and /VC:/ in stressed syllables. This, in turn, means that only native speakers of Norwegian, Icelandic and Italian would, as L2-learners of Swedish, be expected to automatically adjust consonant length after preceding vowel length. The fact that a vast majority of L2-learners of Swedish have L1’s without complementary vowel-consonant length, ought to result in teaching strategies where both vowel and consonant length are taught actively, not least considering the role of the postvocalic consonant to give the proper length to the stressed syllable, as will be further discussed in 1.5.1 through 1.5.3.

According to the “feature hypothesis” (McAllister et al. 1999) a phonologic contrast or category that is based on a phonetic feature that is not utilized in L1, is harder to learn in an L2, than a phonologic contrast or category which is based on a phonetic feature that is utilized in L1. McAllister et al. (1999) compared the ability by speakers of Spanish, English and Estonian to master the Swedish quantity contrast, and they found that the ability corresponded to the degree to which, the L1 utilized duration, which would in this case be the phonetic feature of interest. This means that the Swedish quantity contrast, and the durational aspects of stress, cannot be expected to be learned automatically. Furthermore, Jamieson & Morosan (1986, 1989) found that although learners could profit from hearing exaggerated synthetic examples of a specific contrast, and apply it to natural speech, they could not generalize it to other positions in the word. This implies that even though a learner’s L1 utilizes duration in the realization of a vowel contrast, this learner may encounter difficulties increasing the duration of a postvocalic consonant. One conclusion drawn from this discussion is that the length of the postvocalic consonant in Swedish should be taught actively to all L2

learners of Swedish. The few learners with Norwegian, Icelandic or Italian (cf. above, same subsection) as L1, could then enjoy the ease by which they (probably) learned to master the complementary length pattern of Swedish.

The temporal correlates of the complementary length pattern in central standard Swedish (Stockholm variety) are clear from Elert (1964); the duration of phonologically short vowels divided by the duration of phonologically long vowels yields 0.65. And the duration of phonologically short consonants divided by the duration of phonologically long consonants yields 0.75 (words in sentences) to 0.80 (words in isolation). Teaching vowel and consonant length as equally important could consequently be expected to promote the naturalness of spoken Swedish as an L2, in spite of the difference in short/long-ratios.

Fischer-Jørgensen & Jørgensen (1969) report clear cases of longer postvocalic consonants after short vowel in North German, with C/C:-ratios near those found by Elert in central standard Swedish (Stockholm variety). This is surprising, since standard German otherwise is said to be a language without complementary consonant length. Bavarian, however, is reported to have this relation (Bannert 1976). On the other hand, Gårding et al. (1974) report smaller differences between long and short consonants in South Swedish (Skåne variety); C/C: 0.84. Data recently collected by the present author from the Swedia (Eriksson 2004) material (20 speakers of the Skåne variety), showed the same mean value, but observations ranged from C/C: values 0.3 to 1.3. Malmberg (1944) suggested that the Skåne-variety of quantity be classed as involving only vowel length. Schaeffler (2005) studied the realization of quantity in the whole Swedish speaking community, including Finland, with respect to absolute durations of vowels and consonants, as well as the relative values V:/V, C:/C, V:/C and V/C:, which resulted in three realization types: Finland-Swedish, Northern Swedish and Southern Swedish. The boundary between Northern and Southern Swedish did not separate Skåne from the rest of Sweden, rather is there a boundary at the latitude of Stockholm or a bit further north. These results are however treated with caution here, since they are based on one single sequence of segments, namely /ak/ as in *tak* [tɔ:k] ‘ceiling/roof’ and *tack* [tak:] ‘thanks’. According to Behne et al. (1997), there is an interaction between spectrum and duration in the signaling of /V:C/ and /VC:/ in the case of /a/. The same combination of duration and spectrum cannot be assumed to function similarly over the whole vowel inventory, since e.g. /ɛ/ relies almost entirely on duration for the signaling of quantity category (Hadding-Koch & Abramson 1964, Jonasson & McAllister 1972). This issue will also be addressed in the present work, section 3.1 and 3.2.

These somewhat inconsistent findings indicate that there is a great variation in durations of postvocalic consonants, and some of the variation can be suspected to be caused by the test situation. The findings in North German, compared to standard Swedish and the standard descriptions of the respec-

tive languages cannot be accounted for with available findings. The occurrence of complementary consonant duration in German is not problematic for BP, but the total lack of complementary consonant duration in Swedish varieties would be.

Two contexts with high demands on clarity and distinctiveness are focal accent and infant directed speech. The word carrying the greatest informative load in a sentence, normally receives focal accent, since it is mostly the least predictable word in the sentence, and hence must be given the richest phonetic equipment. These two conditions show greater temporal difference between /V:C/-words and /VC:/-words. In the case of focal accent Heldner & Strangert (2001) found that /V:C/-words received their extra duration distributed over the entire stressed syllable, while /VC:/-words received their extra duration mainly in the postvocalic consonant, and the vowel received no extra duration. This way the temporal difference between /V:C/-words and /VC:/-words is increased in focal position. Sundberg (1999) found that mothers made a general increase of vowel length in infant directed speech compared to adult directed speech, although “a more detailed analysis shows that short vowels are shortened and long vowels are lengthened in the transformation from ADS to IDS. “... Much the same happens with consonants: short consonants are shortened and long consonants are lengthened.” (Sundberg 1999: 2190). The complementary pattern, involving both vowel and consonant in stressed syllables seems to play a role in situations that demand a high degree of clarity or distinctiveness.

One of the points of the present thesis is to utilize the option to lengthen stressed syllables with a short vowel, by means of the complementary consonant length, thus promoting both possible perceptual cues to quantity (see section 4, this thesis) and well known perceptual cues to word stress (Fant & Kruckenberg 1994), as well as global naturalness. In pedagogical descriptions, like e.g. Håkansson & Stenquist (1989) where consonant length is not given a prominent place, one can suspect that the authors have read or heard researchers that claim consonant length to be redundant, i.e. predictable from vowel length, and interpreting “redundant” as not important per se, and paying respect to researchers in phonology, they do not list the consonant length among preferred learning goals e.g. Bannert 1979b, 1980). It should however be pointed out that there are experts in Swedish phonology, e.g. Riad (1992:307) who propose a non-redundant role for the postvocalic consonant length. A couple of modern university text-books in phonetics (Engstrand 2004:183, 2007: 73) do also describe the Swedish quantity in agreement with BP, i.e. as two alternative ways of lengthening a stressed syllable.

Experiments with manipulated duration of postvocalic consonants have been carried out by Hadding-Koch & Abramsson (1964) and by Behne et al. (1998), but they found no effect of changed consonant duration on the quantity categorization made by native Swedish listeners. This issue will be addressed in the present thesis, section 3.2. Results like these, that point out the

postvocalic consonant as a non-distinctive perceptual cue to quantity, adds fuel to the view that the postvocalic consonant has no important role to play in Swedish in general, and hence not in the teaching of Swedish as an L2. This is an unfortunate confusion of the interests of phonology on the one hand, and the interest in teaching Swedish as an L2 on the other. It is stressed here, that apart from rendering naturalness to Swedish speech, the length of the postvocalic consonant is an important perceptual cue to stress, especially in syllables with a short vowel (cf. above, same subsection). The great importance of stress on lexical and sentence level is acknowledged by most people engaged in teaching Swedish as an L2 (personal communication with many teachers over many years). The role of the postvocalic consonant is however not emphasized to the same extent. An example is an updated issue of a manual for teachers and learners of Swedish as an L2 (Ballardini et al. 1997, 2001). All of the chapters are followed by vocabulary, grammar and/or pronunciation exercises. Most of the pronunciation exercises contain examples of stress, mostly various types of sentence stress. In the earlier chapters, for beginners, there are outlines on Swedish long and short vowels, which are marked for length and shortness respectively. An underlined vowel-letter indicates lengthening of a stressed /V:C/-syllable, but nothing but shortness is signaled by the dot under the vowel-letter in stressed /VC:/-syllables. The example below (left) shows how length and shortness are marked in the mentioned work, and to the right is shown a BP-account of the same word pair.

vila - villa

vila - villa

It seems reasonable to include postvocalic consonant length in rules for Swedish pronunciation, since it is obviously part of a durational pattern in spoken Swedish. From a pedagogical perspective, length can be regarded as both a phonological feature and its physical correlate, since this distinction is not assumed to be made in most teaching situations. The term “length” is also assumed to be easily associated with temporal duration by a learner of Swedish as an L2. A system of directives for L2-pronunciation should be consistent, promote cues to important phonologic properties as well as global naturalness. The part that is played by the postvocalic consonant in both the complementary pattern of quantity categories, and the means it provides to increase duration of stressed syllables, makes it a good ingredient in Swedish pedagogic phonology.

The duration of the postvocalic consonant, and the controversy about its role in the field of L2-teaching of Swedish as an L2, has been a driving force behind the present thesis. The postvocalic consonant is a temporal factor that makes L2-students sound so much more “Swedish” in the subjective perception of the present author, and by many teachers (personal communication) with whom he has discussed this issue. The studies in the present thesis will attempt to answer a few questions concerning the consonant length, particu-

larly its role as perceptual cue to quantity and stress. This will be made partly by letting a number of L2- speakers of Swedish, with a great variety of L1's, produce Swedish words that have a corresponding sequence of identical segments but another stress or quantity category. Eventually the temporal pattern in native Swedes' productions of English and German will be examined. It should be understood that "consonant length" is a pedagogic term, used primarily to make learners increase the durations of postvocalic consonants in /VC:/-syllables. The studies examine the interplay between segment durations and native listeners' perception of categories and naturalness. The questions central to the present thesis are listed here:

- Can the postvocalic consonant contribute to the signaling of quantity, in spite of the conclusions from previous experiments (Hadding-Koch & Abramson 1964, Behne et al. 1998)?
- Can an L2-speaker produce acceptable, i.e. unambiguous Swedish words with respect to quantity and word stress, without utilizing the postvocalic consonant duration?
- Is the complementary pattern of Swedish stressed VC-sequences persistent enough to influence L2-speech in native Swedish speakers?

We know rather well from experiments reviewed by Fant & Kruckenberg (1994) that some measure of syllable duration correlates positively with native Swedish listeners' perception of syllable prominence. Fant et al. (1991b) shows clearly how the postvocalic consonant constitutes a greater share of the syllable duration in /VC:/-syllables, compared both to the preceding vowel, and to corresponding consonant duration in /V:C/-syllables. Will the same pattern be observable in unambiguous L2-realizations of Swedish word stress categories, or will L2-speakers manage to signal stress in Swedish by means of non-typical cues? One can also ask how strong the complementary vowel-consonant duration pattern is in native Swedish speakers. Does this pattern appear as native Swedes speak other languages that do not utilize vowel and quantity duration the same way as Swedish? As mentioned above in 1.3.1, Lindblom et al. (1981) claim that this is the case, and it is also an observation made by Engstrand (2007: 77).

## 1.5. Some theoretical and pedagogical considerations

Before we continue, we should recall that the high priority of temporal prosody in Swedish pronunciation is an assumption based mainly on experience and intuition, but that there is some support based on empirical studies. Some of the consequences of both the description and the emphasis on temporal properties being basic for intelligibility can be worth looking at more closely. Below follows a discussion concerning the consequences of the BP-approach and its three rules involving sentence, word, syllable and segment



prosody. The discussion is divided into three aspects; 1) phonologic features, 2) phonetic correlates and 3) pedagogic considerations related to the BP principles. It is noted here that the notions of ‘priority’ and ‘importance’ are not restricted to mere phonetic learning outcome in terms of listener friendliness, but does also include profitability in the teaching situation, an aspect developed further below, in 1.5.3. The theoretical considerations deal mostly with the structure of the Swedish phonologic system, and variations found in its realization. The three aspects cannot be kept totally apart, since the whole idea of priority has a pedagogic background, but attempts are made to focus on aspects relevant to the headings. Also recall from 1.4.1 how the term “length” is used in a sense that does not distinguish the phonologic and acoustic levels.

### 1.5.1. Priority among prosodic phonological features

Since rules of stress, and not rules of the tonal word accent are included in the set of basic pronunciation rules for L2-learners of Swedish, the former are assumed to be more important to most L2-learners of Swedish, in order to achieve a listener friendly speech, than the latter. See table 1.1, column 3, for examples of Swedish word accent.

This assumption is supported by the extent to which the Swedish phonologic system utilizes stress as opposed to word accent. Word stress is realized in more prominence degrees than the tonal word accents. The prominence degrees are shown in table 1.3 below, which is an adaptation of a figure in Bruce (1998: 80). Table 1.3 can be seen as different degrees of phonetic richness, where a higher stage contains all the properties of a lower stage plus an extra property. In the unstressed version, the word is just a string of segments. In the secondary stressed version, the string of segments has a temporal structure, i.e. a monosyllabic word has received extra length in one segment, and a polysyllabic word has one or more short syllables in addition to the one containing a long segment. In the accentuated version, there is a tonal gesture (typically a fall) rendering it possible to be identified as having accent 1 or accent 2<sup>7</sup>. In the highest stage, focal accent, there is an additional tonal gesture (typically a rise), which should tell the listener that this is the most important word of the phrase. Moreover, as mentioned above in 1.5.2, the focal accent entails additional duration in the stressed syllable. This means that the phonologic features that involve duration in their realization are present in all stressed positions, and the features involving tonal realization are added in higher stages of word prominence. This can be interpreted as a means of the linguistic system to protect the temporal features

---

<sup>7</sup> According to Bruce (1977) both accent 1 and 2 have a tonal fall, and according to other authors, e.g. Elert (1966) accent 2 has a specific tonal contour and accent 1 only borrows the tonal contour of the sentence intonation. The relevant aspect here, is that the two categories are possible to distinguish.

more than the tonal ones. The stages mentioned are assumed to correspond to perceived prominence by the listener.

Table 1.3: Manifestation of phonologic category/contrast at different prominence levels

Phonological contrast Main phonetic manifestation	Quantity contrast 1 segment with extra duration	Word accent contrast Falling tone	Rising tone
Unstressed	NO	NO	NO
Secondary stress	YES	NO	NO
Accentuated	YES	YES	NO
Focal accent	YES	YES	YES

Stress is realized more uniformly in the entire Swedish speaking community, than the tonal word accents. Moreover, we can state that the Swedish word accents 1 and 2 are neutralized in standard Finland-Swedish (cf. Gårding 1974a: 67), a variety regarded by most Sweden-Swedes as not difficult to understand. The standard Finland-Swedish dialect has occurred frequently in a popular animated cartoon directed to pre-school age children (and upwards). The present author has never heard reports of difficulty in understanding the spoken Finland-Swedish by Sweden-Swedish listeners. In addition to the absence of word accent contrast in some dialects, the manifestation of accent 1 and 2 are rather different in different regional varieties of Swedish (cf. Gårding 1974a). Some of these differences can be seen as difference in sort, e.g. two-peak accent 2 vs. one-peak accent 2, and some can be seen as difference in degree e.g. timing of tonal gestures (Gårding & Lindblad 1973). There are also regional differences with respect to the distribution of word accent categories, a fact that is incidentally mentioned by Elert (1970) and Garlén (1988). It was striking to the present author, how many words changed from accent 2 to accent 1, when he moved 400 kilometers north from Stockholm. Moreover, if a Swedish text is sung, the word accent contrast is neutralized by the melody of the song. The word accents should be manifested primarily as tonal falls, and in focal position as an additional rise (Bruce 1977), and the tones of songs are more or less stable, i.e. each note, corresponds roughly to one syllable, has a constant pitch, and it does not sound like music – in a traditional sense – if rises and falls are introduced in singing (cf. Gårding 1974a). Great artists like Bob Dylan can allow themselves to do so occasionally, but it is not standard.

Ambrazaitis (2007) found that expected tonal patterns for accent 1 and 2 respectively, were found only in about 60 % of the cases, in a sentence where the test words were assumed to attract focal accent. This can be interpreted as though the realization of word accents plays a minor role for the intelligibility of Swedish (cf. Finland-Swedish above). The system seems to accept that word accents are neutralized in secondary stress position, in dialects and in singing, but neutralization of lexical stress can only be acceptable in unstressed positions. It should be noted that some Swedish songs are not optimally matched with respect to lexical stress and time value, occasionally rendering words ambiguous. A well known example is the hymn *Tryggare kan ingen vara* ‘No one can be more secure’, where *tryggare*, normally a dactyl [trÿg:are] in the melody is given the timing of an anapest, [trygaré:]. The combination with the auxiliary *kan* [kan] ‘can’ i.e. [tryga're:kan]<sup>8</sup>, is acoustically very close to [trygarè:kan], meaning ‘the secure shrimp’ a phrase that does not give a religiously solemn impression.

Attempts have been made to empirically evaluate the relative importance of stress and tone for intelligibility, comprehensibility and naturalness of L2 Swedish. Bannert (1995) made digital corrections of foreign accented Swedish, temporally, tonally, and both methods in combination. The speakers’ L1’s were Persian and Punjabi. The tonal correction showed better effect in terms of perceived intelligibility<sup>9</sup> and acceptability compared to the temporal correction, and tonal combined with temporal correction gave best result. The opinion of the present author is however, that the temporal correction was not made according to the BP-rules presented above. All phonologically long vowels were given increased duration, but no phonologically long consonants were temporally altered, and in one case a phonologically short vowel was given longer duration. Norwegian is quite similar to Swedish with respect to stress, word accent and quantity, and Holm (2006) made digital corrections of foreign accented Norwegian; global temporal organization and global intonation. The effect of temporal and tonal correction respectively, depended on the speakers’ L1’s. Temporal and tonal correction in combination always yielded best results. In both Bannert (1995) and Holm (2006), the tonal correction was global, and did not separate word accent gestures from global sentence or phrase intonation.

There is no study, known to the present author, that reports intelligibility problems with erroneous word accent realization, and a cautious conclusion is that native Swedish listeners can tolerate many kinds of deviations from a classical (Bruce 1977) expected realization of the Swedish word accents, without having problems understanding the speech. The present author wishes in this connection to report observations from a male speaker, to

---

<sup>8</sup> Word accent mark is avoided in this transcription since it will illustrate the result when sung, as word accents are neutralized.

<sup>9</sup> Intelligibility in this experiment would correspond to comprehensibility as defined by Munro & Derwing (1995 and 2005)

whom he has listened for quite a while. He speaks Swedish after having learnt Lappish, Finnish and Norwegian. The four languages, including Swedish were learned in childhood, before the age of eight. This person seems to distribute quite distinct realizations of accent 1 and 2 at random, a property that is not experienced as disturbing, only at times as a bit funny. The same syndrome can be heard from speakers from Tornedalen ‘Torne Valley’ in north Sweden, where the first language of many people is a local variety of Finnish (‘Meänkieli’), and Swedish is learnt at school or in day care centers.

As mentioned in 1.5.1, there is great variation in realizations of tonal features, while the realizations of temporal features are more stable over the Swedish-speaking area. Comparing the tonal gestures in the first syllable of an accent II-word, spoken by a person from Stockholm, to the same word spoken by a person from Gotland or Dalarna, there would be a fall in the former case and a rise in the latter (e.g. Gårding & Lindblad 1973), and the word is still mutually intelligible. This can be interpreted as though the tonal movement – whatever it may be – lends prominence to the syllable. A parallel exchange of direction in the temporal organization would be hard to imagine, but as is made clear in 1.5.3, duration is a main perceptual cue to prominence in Swedish, and it is not reversible as is the case with tone. And last but not least; the tonal gesture of the word accents are superimposed on the stress pattern. The stress pattern is a prerequisite for the distribution and the realization of the Swedish word accents. This state of things, that tonal categories present a great variety in their realization, and that there is a one-way dependence between stress and word accent distribution, is a strong support for the BP approach, which regards temporal structure as a more important pronunciation goal than tonal structure.

In BP the quantity contrast is given priority over the word accent contrast. In addition to what is mentioned above, the priority has to do with experience concerning what phonetic deviations are the most harmful to listener friendliness and intelligibility. Moreover, the realization of stress is readily combined with the realization of quantity. By means of phonologic length, realized as extra duration allocated to the correct segment, the learner is likely to manifest two phonologic contrasts with one and the same expedient.

The assumption that length is stress-induced and only a potential property at the lexical level, results in a Swedish inventory of nine vowel phonemes and a prosodeme of length that constitute the quantity distinction (cf. Elisasson 1985, Elisasson & LaPelle 1973, Garlén 1988). To categorize the Swedish quantity contrast as a temporal one, presupposes that the Swedish vowel inventory is defined as having nine vowel phonemes that are altered by means of a length prosodeme.

It has been suggested, e.g. by Malmberg (1956) that the Swedish vowel inventory consists of 18 phonemes, 9 long and 9 short. This view does not support the notion of stress induced length. Whenever Swedish vowel pho-

nemes are mentioned in the following text, it will be based on a nine vowel inventory, if not otherwise indicated. According to Eliasson (1985), Swedish cannot have 18 vowel phonemes, and vowel length should not be considered as lexical, since word stress can change as a consequence of derivation, and segment length occurs and disappears according to stress allocation. Eliasson (1985:108) gives numerous examples and one of them is given here:

<i>sys'tem</i>	[sys'te:m]	'system'
<i>syste'matisk</i>	[systə'mø:tisk]	'systematic'
<i>systema'tik</i>	[systəmə'ti:k]	'systematics'
<i>systemati'sera</i>	[systəməti'se:ɐ]	'systematize'

This example indicates that length follows word stress, and that spectral properties typical to either long or short vowel category do not lie lexically anchored in each vowel, but are triggered by word stress and the segmental composition of the syllable, resulting in /V:C/ or /VC:/.

The notion of stress-governed segment length is central in the present Basic-Prosody approach, and figure 1.2 illustrates how segment length can be regarded in a model of length-assignment as part of a word retrieval process. It is true that lexical stress, quantity category and word accent, often can be predicted from morphological and phonotactic factors, but in BP, which is a simplified description for a particular pedagogic purpose, the main point is that the prosodic features are seen as potential properties of the word, which may or may not be realized in speech. One can also assume that a learner of Swedish as an L2 does not have the same intuition with respect to lexical, morphological and phonological aspects as a native speaker. If a portion of extra duration is assigned the proper segment of the proper syllable of the proper word, the Swedish speech is assumed to be intelligible and listener friendly, understood that phonetic quality of segments comply with minimal requirements of distinctiveness.

A traditional way of describing the Swedish quantity contrast is to define it as "vowel length", which in turn means that "long or short vowel" are the two quantity categories. This view can be exemplified by Håkansson & Stenquist (1989) and Ballardini et al. (1997, 2001), which were discussed in 1.4.3 above. In these teaching media, an underlining or a dot indicates vowel length or vowel shortness in stressed syllables. The complementary relationship between vowel and consonant length is by no means denied by this description, but it does not emphasize the consonant length as anything worth mentioning in an L2-learning context. An older manual (Higelin et al. 1973) encourages in the teacher's guide, teachers to draw the learner's attention to prosody by marking stress and length in texts. In the example given, there are main stress apostrophes on words in focal position and underlined

vowels in all words of higher prominence than secondary stress. No marks for short vowels or long consonants. In the BP-description, the quantity contrast is seen as a phenomenon involving vowel and consonant duration with equal importance.

BP presupposes that all segments are “short” if they are not given the extra duration that comes with the stress. As a consequence of regarding phonological segment length as stress governed, the quantity distinction is not seen as primarily a vowel feature, but rather as two ways of lengthening a stressed syllable – two ways that make up a phonologic distinction. The BP-approach wishes to kill two birds with one stone, by utilizing the fact that i) a long postvocalic consonant helps increase the duration of a stressed syllable with a short vowel, a syllable that wouldn’t have been lengthened if the phonological rules had only accounted for vowel length, and ii) the complementary pattern is prevalent and can be assumed to render naturalness to the pronunciation.

### 1.5.2. Priority among phonetic correlates

Since only length is given in the BP-rules, as a means of realizing stress in general, it is assumed that duration is the most important physical correlate of stress, more important than intensity, tone or spectrum. Can this assumption find support?

There has been an intuitive assumption that perceived stress or prominence in Swedish was associated with higher intensity, e.g. “expiratory accent” in Elert (1966: 86, Thorén 1988: 6). The empirical support for this view is however limited. Fry (1955) studied stress in English disyllabic words, and found that vowel duration ratio was “a more effective cue than intensity ratio”. Fant & Kruckenberg (1994) found intensity to “play a subordinate role for Swedish, comparing nonfocal stressed and unstressed syllables”. Moreover, they conclude from earlier studies:

“duration is the most consistent physical correlate of stress. Next, or of equal importance, is the F0-pattern of the associated word accent. However, we find examples of weak stress realized by a significant duration increase but without F0 modulation.” (Fant & Kruckenberg 1994: 141)

The second sentence in this latest quote can be compared to table 1.3, where secondary stress is not expected to entail any specific tonal gestures in its realization, only the stress induced duration. The assertion that stressed syllables have longer duration than unstressed ones in Swedish, is also supported by Strangert (1985), Eriksson (1991) and Thorén (1982), and compellingly explained by Aurén (1869), who labeled sentence stress and word stress as “word quantity” and “syllable quantity” respectively, thereby signaling that some aspect of ‘length’ is the manifestation of the phonological contrasts. Figure 1.4 from Fant & Kruckenberg (1994) illustrates how syllable

ble duration is dependent on both stress and number of syllables. The figure also shows the interesting relation, that a distinct reading mode does not result in an overall lower speaking rate, but increases the durational difference between stressed and unstressed syllables. This in turn, implies that greater distinctiveness, or clearness, in Swedish is not primarily making every word more prominent, rather the distinctiveness is realized as a greater temporal contrast between stressed and unstressed syllables.

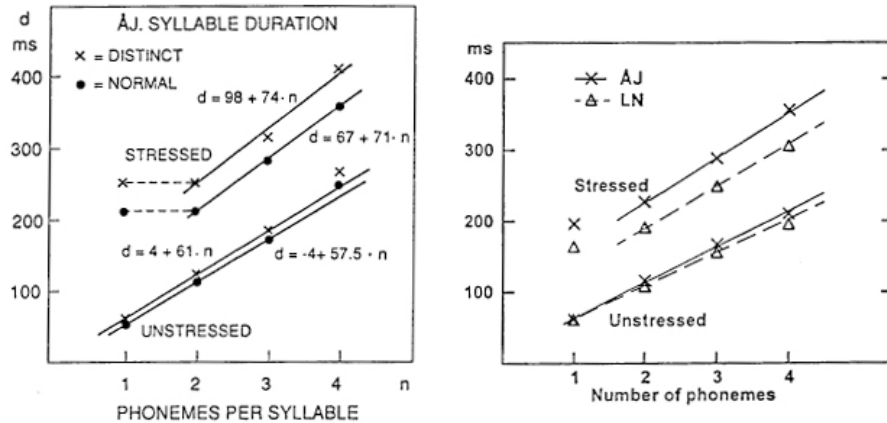


Figure. 1.4. Duration of stressed and unstressed syllables as a function of the number of phonemes. To the left, subject AJ in normal and distinct reading mode; to the right, subjects AJ and LN normal reading (from Fant & Kruckenberg 1994, with permission).

In the cited article by Fant & Kruckenberg (1994), it is also said that tonal correlates to stress are possibly equal to temporal correlates with respect to how well they correlate with perceived degree of prominence. We are however reminded from 1.5.1 and table 1.3 above, that the temporal structure of lexical stress should be realized in all degrees of stress, i.e. secondary stress, accentuation and focal accent. The tonal contrast of word accent (accent 1 and 2, acute and grave respectively) should be realized only in the two highest degrees among stressed categories, namely accentuation and focal accent. This implies that hypothetically 1/3 of all stressed syllables could be realized without any specific tonal gesture.

Since no other way than length is given in the BP-rules, as a means of realizing stress in general, the duration of the postvocalic consonant is required as a means for realizing stress in a syllable containing a short vowel, irrespective of the possible importance the postvocalic consonant may have as a cue to the quantity contrast. This issue has been touched upon above in 1.5.2 and is a central property in BP. It could not be utilized in e.g. Finnish, where quantity and word stress are distributed independently of one another. Behne et al. (1998) speculated that the duration of the postvocalic consonant is "a temporal artifact of the preceding vowel quantity", and "...that listeners do

not use the duration of the postvocalic consonant to identify vowel quantity in Swedish.” The notion of “temporal artifact” could be taken as a pretext to the BP-claim, that the complementary pattern is useful in an L2-learning context; to lend the expected duration to stressed syllables. Heldner and Strangert (2001) found that a word in focal position had longer duration than words in nonfocal position, and that the extra duration occurred within the stressed syllable. When they measured segments within the stressed syllable, they found that the phonologically long segment, whether vowel or consonant, was lengthened most. They conclude that “the contrast between long and short vowel in stressed syllables was sharpened in focus” (Heldner and Strangert 2001:329). Their findings are in agreement with an earlier study by Bannert (1979a). The postvocalic consonant is thus not only filling in the “missing” duration after a short vowel, but also gives the extra duration to a /VC:/ syllable in focal position, compared to the same syllable in non-focal position.

If we assume that focal accent is a means of giving higher distinctiveness to a word, it can be compared to the speaking styles “distinct” and “normal” in figure 1.5. Both findings can be seen as tokens of increased temporal contrast; in the case of focal accent, we see that the quantity contrast receives extra durational difference between its categories, while a “distinct” speaking style receives greater durational differences between the categories “stressed” and “unstressed”. That is to say that extra clearness in the Swedish language does not result in increased audibility in all segments, but rather in increased difference between phonologic categories, with durational differences occurring in all degrees of stress, and tonal gestures added in higher degrees of stress.

Although we feel sure that duration is a reliable – possibly the most reliable – correlate to stress in Swedish, there are other known correlates. Engstrand (1988) found that the vowels /i u a/ were articulated with narrower constriction in stressed position than in unstressed. Sluijter and van Heuven (1996) found that duration and high frequency emphasis (spectral tilt) were parameters that could distinguish stressed syllables from unstressed syllables in Dutch and English, whereas  $f_0$ -movements and overall intensity were better in distinguishing syllables in accentuated position from syllables in focal position. These, and similar findings made by Marasek (1996), with respect to German, makes it likely that these spectral correlates may be valid also for Swedish, as they are all Germanic languages. We are consequently dealing with temporal (duration), tonal ( $f_0$ ), spectral (voice source and articulation) and dynamic (intensity) correlates that probably interact in the role of perceptual cues to word and syllable prominence. The experience of the present author – gathered during a quarter of a century of teaching – is that learners sound well enough, if they manage to lengthen the correct segments, i.e. it seems like spectral, and to some extent tonal, properties adjust



themselves more or less automatically, when the learner deliberately lengthens a segment, as was discussed in 1.1.

As discussed above in 1.1 and 1.2, the L2 speech of a learner who started as adult, is mostly marked with various phonetic deviations, known as foreign accent. One aspect of foreign accent is the realization of an L2 category by means of other phonetic correlates than what is typical for native speakers of the target language. A non-typical realization of an L2-category may or may not be perceived as the intended category by the L1-listener. A great variation of L2-realizations of Swedish word stress and quantity, could give a hint as to whether temporal correlates are necessary for unambiguous realizations, or not. A number of L2-speakers with many different L1's, and with many different levels of their mastery of spoken Swedish, can be expected to create the desired variation. Two such studies will be reported in sections 3.3 and 3.4.

Since only length – meant to be interpreted as duration – and not timbre is described in the BP-rules, as a means of realizing quantity, duration is assumed to be more important to the perception of the quantity contrast, than spectral differences between long and short vowels. In addition to experience, this statement relies mainly on 1) the fact that spectral differences between long and short vowel allophone is realized differently in different varieties of Swedish 2) Findings from Hadding-Koch & Abramson (1964) indicate that native Swedish listeners perceive words with original long vowel quantity as short vowel quantity after manipulation of only vowel duration, and Behne et al. (1997) successively increased the duration of originally short vowels resulting in native Swedish listeners perceiving words with original short vowel quantity as having long vowel quantity. There is however still uncertainty with respect to the status of the /ɘ/-phoneme in relation to the quantity contrast. It seems from Hadding-Koch & Abramson (1964) that this phoneme differs from the rest of the Swedish vowel inventory, since durational manipulation did not change listeners' categorization of original *ful* [fɘ:l] 'ugly' and *full* [fel:] 'full/drunken'. The authors' conclusion was that /ɘ/ is not sensitive to duration for quantity categorization, and should consequently be regarded as two phonemes; /ɘ/ and /ø/, differing in spectral properties. The fact that the vowel duration in words like *ful* and *full* behave like other long-short pairs, does not support this view, and it is pointed out in Hadding-Koch & Abramson (1964) that speakers and listeners in their study were native in the South Swedish dialect of Skåne, that is known to have somewhat different durational patterns than most varieties in the rest of the Swedish speaking community (cf. Gårding et al. 1974, section 1.4.3.). The method employed by Hadding-Koch & Abramson (1964), i.e. changing the duration of either vowel or consonant, leaving the other segment intact, creates stressed syllables with "un-Swedish" duration, in that a stop closure that is lengthened after a long vowel yields a /V:C:/-syllable, and consequently an originally long vowel that is shortened

before a short consonant will yield a /VC/-syllable, i.e. either an overlong syllable or a short syllable. The BP-view, that the quantity distinction involves the entire VC-sequence in the complementary way described in 1.4.2 and 1.4.3, does not make a strong claim that it is the combination of vowel and consonant duration that is the main cue to the listener's categorization, but it leaves a gap in the examination of the nature of Swedish quantity. More sophisticated was the study by Behne et al. (1998), where consonant duration was altered, while the vowel was held constant midway between typical values for long and short allophone, with respect to both duration and spectrum. The pretext was that the vowel should thus be neutralized and not affect the categorization of the listeners, and the effect of postvocalic consonant duration could be evaluated. We must however consider that either vowel or consonant duration is likely to be perceived relative to other entities, partly because quantity categorization is made at different speaking rates. So whatever duration we assign the vowel, in order to examine the effect of consonant duration, the vowel duration will contribute to the context to which the consonant duration is related. The issue of relativeness will be discussed further in 1.6. A study that fully recognizes the interplay between vowel and consonant duration, will be carried out in the present thesis, section 3.1. Vowel and consonant durations will be manipulated in a complementary way, i.e. increase in vowel duration is accompanied by decrease in consonant duration and vice versa, thus maintaining the total duration of the VC-sequence just by displacing the boundary between the segments of interest (cf. below 1.5.3, figure 1.8, and section 3.1.2).

It is true that Swedish long and short vowel allophones show spectral differences, and the phonemes /a/ and /ʌ/ are known for great spectral distance between long [a:] and [a], and [ʌ:] and [ə] respectively. The rest of the vowel phonemes show moderate to negligible spectral differences between long and short allophones, although all Swedish vowel phonemes but one, namely /ɛ/, are usually transcribed with different IPA-characters for long and short allophones. There are however Swedish dialects that make no or little, or other kinds of spectral differences between long and short allophones: Standard Finland-Swedish uses [a] for both long and short allophone, and [ʌ] for both allophones of /ʌ/. (Reuter 1971). Some southwest Swedish dialects uses [ɐ] for both long and short allophone of /a/, and among groups of speakers in the South-Swedish area between Helsingborg and Lund, [ʌ] is used for both long and short allophone (Ingers 1970:43). From Elert (1964) it is clear that all Swedish vowel phonemes, including /ʌ/, show substantial durational differences between long and short allophone, and this indicates that there is no strict trading relation between duration and spectrum in the realization of quantity categories, although there are tendencies towards such a relation. There is rather a durational manifestation always present, and a spectral manifestation varying between vowel phonemes and geographic regions. Considering the situation for temporal and spectral correlates in the

whole Swedish speaking society, BP concludes that durational difference between long and short vowel allophones is a more stable correlate than the spectral differences. The greater stability in the temporal aspects of the quantity contrast throughout the Swedish speaking community with all its variation, speaks in favor of BP. Table 1.4 shows the ratios of short vowel allophone to long vowel allophone. Ratios are expressed as percentage, and a low value means great temporal difference between long and short vowel allophone and vice versa.

Table 1.4. The ratio of short allophone over long allophone (V/V:), expressed as percentage for different vowel phonemes. Mean values (From Elert 1964: 113. With permission).

	i	e	ä	y	u	ö	o	å	a
One-Word List (V/t/)	61	60	63	59	83	59	81	59	65
Sentence List (V/t/ + V/s/ only)	70	66	62	64	77	66	69	65	69

The Swedish quantity distinction does, as mentioned above, show durational as well as spectral differences between the long and short vowel category. If we conclude from the reasoning above, that durational correlates to the Swedish quantity distinction show more similarities between vowel phonemes and dialects than do spectral correlates, it could be considered adequate to describe temporal properties in a set of basic rules.

A central issue in the BP-approach is the trading relation between vowel and consonant duration in /V:C/ and /VC:/ syllables. The experiments carried out by Hadding-Koch & Abramson (1964), and Behne et al. (1998) do not show that the duration of the postvocalic consonant would function as a perceptual cue to quantity categorization. Results from Traunmüller & Bigestans (1988), indicate however that, although vowel duration related to utterance duration was strongly correlated to native Swedish listeners' category boundary between /V:C/ and /VC:/, the consonant duration related to utterance duration, was also correlated to the same category boundary, although somewhat weaker. The role of the postvocalic consonant as perceptual cue to the quantity contrast will be addressed in a perception experiment in section 3.2.

Another issue concerning postvocalic consonant duration is when a short vowel is followed by two or more different consonant phonemes. According to BP-rule No 3, there is only one segment that should be lengthened, namely the first consonant segment following the short vowel. As pointed out above and shown in figure 1.5, the duration of syllables is strongly correlated to the number of segments they contain, and this opens the possibility

that a /VCC(C)(C)/ syllable could receive its extra duration from all consonant segments that could be considered belonging to the syllable, and that one of them would need to be regarded as phonologically long. There is however empirical evidence from Behne & Czigler (1995), that, when neutralizing the effects of inherent durations, the first consonant following the vowel, receives the largest portion of the duration, although it is shorter than a single long consonant following the short vowel. Their results also show that a short vowel followed by a cluster is somewhat shorter than a short vowel followed by a single long consonant. The consonants and clusters examined by Behne & Czigler (1995), were /k/ /k:/ /k:s/ and /s:k/. The mean durations of vowel and coda durations in Behne & Czigler (1995) are shown in figure 1.5. The extra duration assigned to the first consonant of a cluster is also supported by the findings of Fant et al. (1991b).

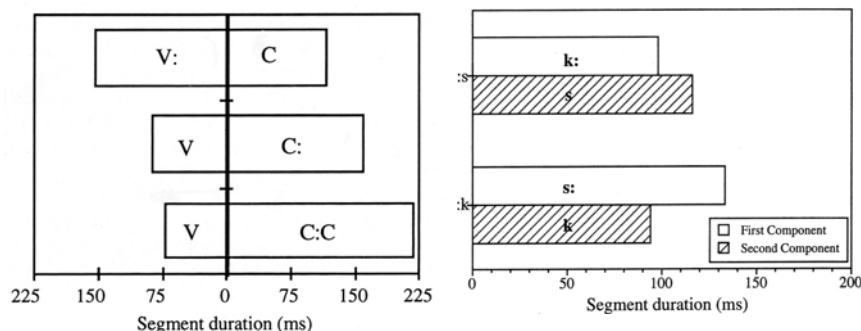


Figure 1.5. Left panel: Aligned mean vowel and coda durations in V:C, VC: and VC:C syllable rhymes. Right panel: Mean durations of the first and second components of the consonant clusters /k:s/ and /s:k/. From Behne & Czigler (1995) with permission.

### 1.5.3. Pedagogical considerations

#### 1.5.3.1. The rules

The BP-approach attempts to be useful. Features or rules that are included in BP should: i) contribute to listener friendliness (if acquired) ii) not be expected to be acquired easily without instruction, and iii) be possible to acquire.

- BP-rule No. 1 says: Stress the correct/suitable word(s) in the sentence.

This rule implies that it is not necessary to stress all words in a sentence. The learner is supposed to adopt this rule either by being exposed to a lot of examples, or by getting it explained in his/her L1. It is thereby not said that sentence stress is introduced in the teaching/learning situation prior to word

stress and segment length. The learner could however believe that all words carried equal stress, if nothing but word stress and segment length were mentioned. Demonstrating sentence stress requires sequences of at least two words. One difficulty worth practicing could be articulating a number of unstressed, short syllables, rather quickly without “stumbling”.

- BP-rule No. 2 says: Stress the correct syllable(s) in the stressed word(s) by making these syllables longer than surrounding syllables.

This rule implies that the correct syllable to stress is not automatically found by counting from the beginning or end of the word, as is the case in e.g. Finnish (always the first syllable in a word) or Polish (always the penultimate syllable in a word). The relevance of this rule can be discussed, since it does not tell the learner how to realize the lengthening of the stressed syllable. This is however specified in rule 3. The rules could go directly from finding stressed word(s) in a sentence, to lengthen the correct segment(s) in that/those word(s), skipping the notion of syllable, which is an entity rather hard to define. The philosophy in BP is however to assume, that the learner can profit a lot from having an idea – albeit fuzzy – of the syllable level, especially when considering the relationship between e.g. iambic, trochaic, dactylic etc. words. In those cases the syllable level can show similarities between groups of words that show differences in e.g. quantity category. Many learners of Swedish as an L2 are also assumed to be familiar with the syllable concept, and can probably profit from descriptions of Swedish, containing the notion of syllables.

At the same time as the present author advocates the priority of temporal properties over tonal, it is admitted that a totally monotonic speech would be all but listener friendly. The risk that any learner of Swedish would sound like a priest singing the mass or service in one tone is however not considered imminent. Moreover, an acceptable tonal gesture should be statistically easier to perform without specific instruction, since there are in the Swedish linguistic community, so many different ways of realizing (and distributing) the tonal word accents (cf. Gårding & Lindblad 1973 and the discussion in 1.5.1.).

- BP-rule No. 3 says: Lengthen the correct segment (vowel or immediately following consonant) in the stressed syllable.

The first virtue of this rule is to emphasize the view that the quantity contrast is not associated only to the properties of the vowel, but rather to the VC-sequence. Another point is that the rule mentions length and not shortness. The priority of temporal features in BP can also profit from the findings of Cebrian (2007), Flege et al. (1997), Escudero & Boersma (2004), Wang & Munro (1999), that show a strong tendency for L2-learners of English, who

have no quantity or tense-lax distinction in their L1, to use duration to a higher degree than spectrum, as a perceptual cue to the English tense-lax distinction. This could possibly be interpreted as if the temporal properties pertaining to the quantity contrast would easily be acquired automatically, but the experience of the present author speaks against this interpretation. It is however advantageous to the field of Swedish as an L2, that the previously mentioned inclination to rely mostly on duration when they perceive, and hopefully later produce, the Swedish quantity distinction, since the BP-approach gives higher priority to duration in this respect.

The wording of rule 3 means that the extra length in the stressed syllable always lies in one segment and is not distributed over a number of segments. Increasing the duration of a vowel is considered rather unproblematic. Increasing the duration of consonants requires at least two main strategies for the realization; i) sustain a nasal, a liquid or a fricative, ii) managing the principal three steps of realizing a stop consonant: initiating a closure during the finale phase of a preceding vowel, sustaining the closure for a longer period than most learners are used to, and eventually releasing the pressure in the explosion phase. The extra, stress induced duration of a consonant cluster following a short vowel, is assumed to lie mainly in the first consonant of the cluster, which is discussed above in 1.5.2. A rule stating that the total length of a stressed syllable is concentrated to one segment, is simpler than one stating that the length sometimes is in a single segment and sometimes is distributed over a sequence of segments.

#### **1.5.3.2. Robustness and elasticity**

Since a teaching situation would hardly contain instructions like: “could you please increase your V/C-ratio to at least 1.5”, or “please make the /t/-occlusion 50 milliseconds longer”, it is advantageous if our rules are robust, which in this case means that realizations of intended pronunciation goals can be exaggerated without violating phonological boundaries or hazarding the listener friendliness. In this reasoning, we assume that BP contains features suitable to exaggerate. There is sometimes a need to exaggerate examples in order to help students perceive a feature or a contrast they are not used to, but the present reasoning is more concerned with the learning outcome, i.e. how the learner eventually speaks as a result of the education. To illustrate the reasoning, we can picture an L2-learner of Swedish who is aiming at pronouncing the word *kal* [kɒ:l] ‘bare/bald’. Since there is another word *kall* [kal:] ‘cold’ containing the same sequence of phonemes, the student should strive to, or be exhorted to pronounce the word in an unambiguous way. This can be done in mainly two ways 1) the temporal way: make the /a/ very long and do not dwell on the /l/, 2) the spectral way: make the /a/ very “dark”, understood as [ɒ]-like, and [a]-unlike. What happens if the student manages too well, i.e. takes phonetic variables beyond an optimal value? In case 1) it could sound exaggerated and a bit funny, and in case b)

the word would probably be perceived as *kål* [ko:l] ‘cabbage’, meaning that a phoneme boundary has been transgressed. The possibility to exaggerate certain features without jeopardizing phonologic boundaries can be called “phonetic elasticity” and the properties we include in our basic rules should be robust enough to tolerate a variety of degrees in their realization, as long as e.g. the duration of a segment is above a critical value. The elasticity is for natural reasons more suitable for temporal than for spectral goals. This means that the vowel of a /V:C/ -word could attain a great range of values, provided that they are all above a critical value for the quantity category. The same goes for the postvocalic consonant of a /VC:/-word, although the duration of the consonant is hitherto not shown to serve as a perceptual cue to the quantity categorization. See however the experiment reported in section 3.2. Consequently a phonologically short segment has a maximal relative duration, in order not to hazard the quantity categorization. The duration of the consonant is anyway expected to contribute to the proper duration of the stressed syllable in order to signal the rhythmical structure of the word, i.e. tell the listener whether the word is trochaic, iambic, etc.

As touched upon above, and which indicates partly divergent goals for research and pedagogy respectively, linguistic instruction typically contains structures that are not identical to typical natural L1 use (cf. the discussion in 1.1.). In a classroom situation, simplified structures and exaggerations are common, which means that the teacher must have an idea of what is worth exaggerating. Jamieson & Morosan (1986, 1989) showed that learners could profit from hearing exaggerated synthetic examples of a specific contrast. The ability to identify categories pertaining to the contrast was generalized to natural tokens and to different voices, and a cautious conclusion is that spoken exaggerated examples could have the same effect. In the mentioned study, the effect was however not generalized to other positions in the word. We hypothesize that the “perceptual fading technique” (Jamieson & Morosan 1986), is possible to generalize to prosodic contrasts. So far, so good, but the choice what to exaggerate is crucial, both with respect to what phonologic features and the way they are phonetically realized. In the case of Jamieson & Morosan (1986, 1989), there was an increased duration of the target segments, although it was the articulation that was of interest, and not the duration per se. In the case of stress in Swedish, and in the light of the discussion in 1.5.2, it seems logical to exaggerate the durational difference between stressed and unstressed syllables, by means of increasing the duration of the phonologically long segment in stressed syllables and shortening all other segments, thus promoting the quantity distinction as well as the word stress distinction. There is a principal difference between the exaggeration of a contrast and the exaggeration of duration. Applying the attempt to exaggerate the contrast between long-short syllables or long-short segments, entails a larger space for lengthening of long segments than shortening short segments, since shortness has a natural terminal point, i.e. when duration

approaches zero and the segment is no more audible. Phonologically long segments will then be the naturally suitable subjects of great exaggerations. This is an example of phonetic elasticity, i.e. a physical possibility to increase the duration of a segment. In a BP-perspective it would be interesting to learn how much of this physical elasticity is present in natural spoken Swedish.

So hypothetically, a phonologically long segment can attain infinite duration, although this would encounter some practical problems. Recall from Heldner & Strangert (2001) and Sundberg (1999) above (1.4.3.), that phonologically long segments are a category that can receive extra length when extra distinctness or clarity is required. Another instance of long segments with additional duration is an impression of the present author that disfluencies in native spoken Swedish, are often realized by dwelling on a phonologically long segment, vowel or consonant, making this segment many times longer than in fluent speech. Those are some indications of phonetic elasticity, a phenomenon that BP implicitly wants to benefit from. Figure 1.6 presents a small sample from Elert (1964), which provides a preliminary test of

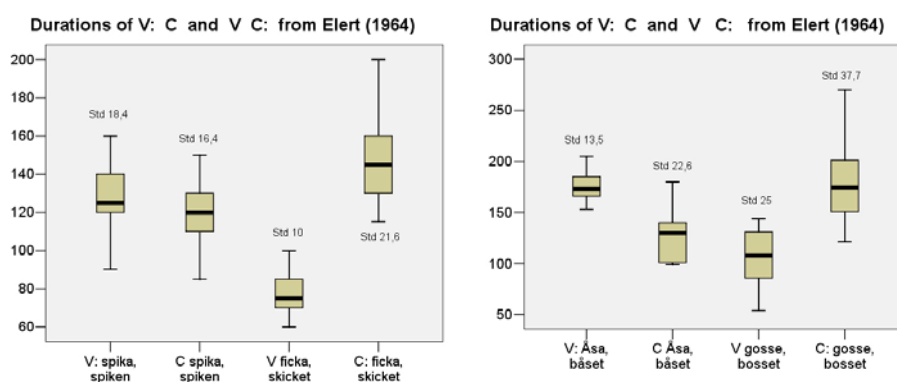


Figure 1.6. Segment durations and standard deviations for four /V:C/- and four /VC:/-words. Number of speakers for all words is 18.

this expected elasticity. Figure 1.6 could be interpreted so, that the elasticity referred to, is present in the case of the postvocalic long consonant in /VC:/-words (the rightmost box in both panels), but not in the long vowel of /V:C/-words (leftmost box in both panels). It can also imply that consonant duration in both quantity categories can attain a greater range of values, due to its assumed lower weight as perceptual cue to the quantity distinction. The figure does not; however tell us that a greater variation upwards for the long vowels, or long consonants, would harm the naturalness or listener friendliness of the pronunciation. This issue is not directly addressed in the present thesis, but the material in most of the present studies will be examined in pursuit of evidence supporting or disproving the occurrence of the phonetic elasticity in L1 or L2 Swedish. Studies 3 and 4 (sections 3.3 and 3.4) provide



some material for this purpose, since they entail a lot of variation in the realization of stress and quantity. A study is outlined, but not yet carried out, by the present author, where recordings of natural L1-Swedish words are manipulated, rendering phonological long segments successively greater durations. The naturalness of the resulting versions will be evaluated by native Swedish judges with respect to naturalness.

### 1.5.3.3. Simplification and adaptation of analysis

Long consonants are often, e.g. in Italian, analyzed as “...the sequence of identical adjacent segments of a sound in a single morpheme ... Because of the syllable division, a geminate sequence cannot be regarded as simply a ‘long’ consonant” (Crystal 2003: 196). Swedish long consonants are analyzed as geminate e.g. in Eliasson & LaPelle (1973). BP however regards Swedish long consonants as simply ‘long’, and the concept and realization of long segments in contrast to short segments, is more important than how long and short segments are analyzed; as geminated, in morae<sup>10</sup>, or close/loose contact etc. The aim of a simplified prosodic description that emphasizes a small number of pronunciation goals is to make it easy for teachers and learners to understand and apply them, without expertise in phonetics or phonology. That is to say that phonologic descriptions in the educational setting should be kept scanty, and expressed in ways that are easily associated with their realizations. The L2-learner must acquire the articulatory techniques to increase the durations of vowels, sonorant consonants, fricatives and stops. To consider whether a phonologically long consonant should be analyzed as one long unit or the doubling of two short units, or a half-long unit plus a short unit, probably does not contribute to the actual pronunciation in the learner. According to the experience of the present author, too many L2-learners of Swedish try to realize phonologically long consonants, especially stops, by producing two explosions. The analysis of long consonant segments as the combination (gemination) of one half-long and one short segment, is theoretically attractive since the ideal syllable has a consonantal onset, at the same time as part of the postvocalic consonant duration is required to increase the duration of the preceding (stressed) syllable. The splitting of the long consonant in two is however in the BP-approach assumed to create more confusion than clearness. The same goes for the mora-analysis. In phonologic theory the mora-analysis of Swedish quantity categories is also attractive, since it preserves the leading role of the vowel as the main perceptual cue to the quantity contrast. Figure 1.7 below shows how morae are distributed in words of the quantity categories /V:C/ and /VC:/ according to Bruce (1998)

Figure 1.8 below, illustrates how a mora analysis could turn out if such an analysis was made in a BP spirit. It would not create a more correct linguistic analysis, but it would perhaps help L2 learners of Swedish to 1) lengthen

---

<sup>10</sup> A mora is “...a minimal unit of metrical time or weight...” (Crystal 2003: 299)

a segment more than they believe possible, and 2) remove the length from the vowel and put “more length than they think” on the consonant. Experience tells us that learners as a group seem to have greater difficulty lengthe-

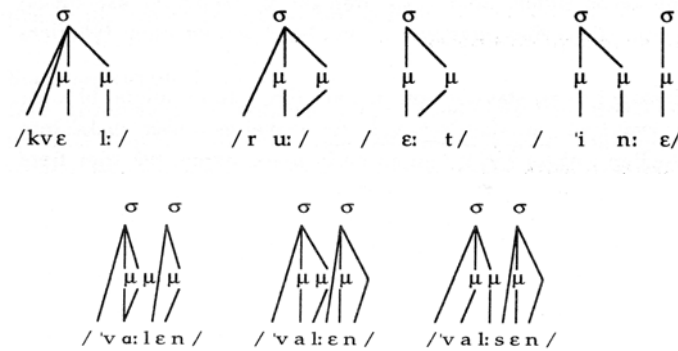


Figure 1.7. Mora analysis of Swedish quantity categories in 7 different contexts. (From Bruce 1998 with permission).

ing consonants than lengthening vowels, and this agrees with the results from Thorén (2006), where a native Polish speaker produced nonsense words that were possible Swedish /V:C/ and /VC:/ words, with native-like vowel durations but without the complementary consonant duration after a short vowel. In this sense BP gives priority to a description that attempts to surmount known difficulties.



Figure 1.8. Mora-analysis of Swedish disyllabic words according to the pedagogical BP-description of Swedish prosody.

A description of the quantity phenomenon, at least the Swedish kind, is that the VC-sequence in the stressed syllable is a production unit that, phonologically speaking, has a constant length, and the difference between a /V:C/ and a /VC:/ sequence is where along this unit that the consonant succeeds the vowel (Bannert 1979b). This is shown schematically in figure 1.9.

V:		C
V	C:	

Figure 1.9. Formalized length relations within the VC-production unit.

To keep the number of rules low, and simultaneously following a strong impression from teaching practice, we assume that length and not shortness must be taught actively; the risk for phonologically short segments to be pronounced long by L2-learners of Swedish is small. This is supported by experience, and is a default rule to avoid too many special rules. BP assumes that “shortness” is the default value for segments in most L1's, and that this value does not have to be taught specifically. According to the present author's experience, this assumption holds for most L1's except for some contexts in Hungarian, Russian and English. In the Swedish of Hungarian learners, all syllables tend to be half long independent of stress. Russian and English learners of Swedish tend to have difficulties shortening a stressed vowel in a /VC:/-word, and giving enough length to the postvocalic consonant. This indicates that shortness sometimes has to be taught explicitly, although the BP-rules only mention length.

## 1.6. Methodological aspects of duration measurements

It should be clear by now, that the present thesis deals mainly with temporal properties of language, and more specifically the duration of segments and sequences of segments. Modern digital technology has made it easy for anyone with moderate computer skills to download free software for speech analysis, e.g. Praat (Boersma & Weenink 2001) and do recordings, measurements and manipulations. This means that measuring vowel and consonant durations is very easy per se (See however 2.1.). If one, however, is to measure segment durations and relate them to the perception of human language users, one is confronted with factors like speech rate, final lengthening (Klatt 1975) and different inherent duration of speech sounds. It is obvious from e.g. Fant et al. (1991a) that the segments of speech have longer duration in slow and distinct speaking mode, than in normal and fast modes. Mean values from 547 segments range from “faster mode” 70 milliseconds, “normal” 75 ms, “slower” 78 ms, and “distinct” 89 ms (Fant et al. 1991a: 253). A conclusion from this is that one can not claim that a long vowel is e.g. 100 ms and a short vowel is e.g. 60 ms. In addition to speaking rate, there is the mentioned phenomenon of final lengthening, that can render a phrase-final or utterance final segment an increased duration, that is measurable, but not perceived by the listener as having longer duration, although the same increase in duration of e.g. a vowel, in another context, probably have influenced the listener's perception of quantity category (see Jonasson & McAllister 1972 below). There is also a well-documented difference in inherent duration in the different vowel phonemes, which is illustrated by figure 1.10. The figure also illustrates how the inherent vowel durations are mainly related to openness, which is in agreement with the “Extent of Movement Hypothesis” (Fischer-Jørgensen 1964). A greater excursion

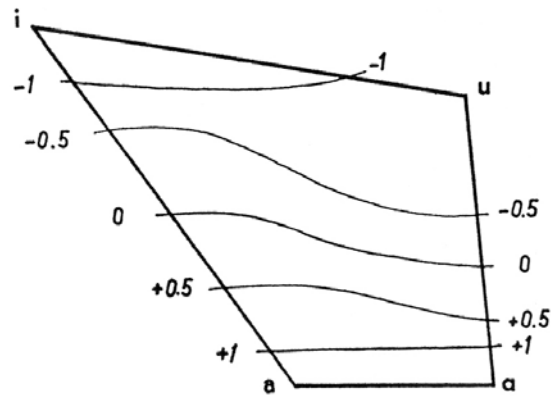


Figure 1.10. Lines marking places in the acoustic-articulatory vowel quadrilateral having the same intrinsic duration value. From Elert (1964) with permission.

downwards of the mandible is assumed to consume more time than smaller degrees of opening, hence the greater inherent duration of open vowels. From Elert's (1964) material it is however clear that the durational difference between phonemic long and short vowel is of a greater magnitude than differences caused by intrinsic duration or by neighboring segments; a long vowel is roughly 55% longer than its short allophone, while comparisons between closed and open vowels show differences of less than 10% (Elert 1964:127). In the light of the discussion above, we realize that what causes a Swedish L1-user to perceive a /V:C/- or a /VC:/-category, must be vowel duration (we disregard for the occasion the influence from vowel spectrum) related to some temporal unit outside the vowel, and since the minimal sequence capable of containing and signaling the quantity distinction is a monosyllabic word, e.g. *al* – *all* [v:l – al:] 'alder – all', *Ohm* – *om* [o:m – ɔm:] 'ohm (electr.) – if/whether' *ås* – *oss* [o:s – ɔs:] 'ridge – us' *är* – *ärr* [æ:r – ær:] 'is – scar', the *contrast* needs at least vowel plus postvocalic consonant to be signaled. Long vowel *category* can however be realized in words ending in a vowel, like *ö* [ø:] 'island' and *bra* [brɔ:] 'good', and exceptional cases of short vowel category signaled without postvocalic consonant can be exemplified by *va?* [va] 'what?'. Jonasson & McAllister (1972) found that vowel duration divided by consonant duration (henceforth V/C-ratio) served as a perceptual cue to native Swedes, in categorizing /V:C/- or /VC:/ in the case of the monosyllabic word pair *ät* – *ett* [ɛ:t- ɛt:] 'eat – one'. They manipulated the durations of vowel and consonant, with maintained total duration of the VC-sequence, which caused the native Swedish listeners to perceive /ɛ:t/ at high V/C-ratios and /ɛt:/ at low V/C-ratios. Although this is true for monosyllabic words presented in isolation, the more natural setting for speech perception is fluent speech, containing phrases and sentences, and a normal unit without pauses could consist of 1-10 words. Diehl & Klunder (1987) claim that: "It is generally not the case that relatively localized

(e.g. syllable-sized) portions of the acoustic signal contain sufficient information to specify phonetic categories unambiguously.” They claim further that: “Experienced listeners make use of *all* potentially relevant cues for phonetic categories, provided these cues are detectable.” These quotes imply that the listener could be expected to relate e.g. vowel duration to at least the duration of a whole word, and possibly to durations of phrases, words and segments in the whole utterance. This is supported by the result of Traunmüller & Bigestans (1988), who found that vowel duration related to the duration of the whole utterance – in this case a 6 word phrase – was the measure that best matched the perceptual boundary between the quantity categories /V:C/ – /VC:/ for the native Swedish listeners. This issue is the basis of the third study in the present thesis, section 3.3, where duration data are collected from a variety of L2-realizations of Swedish words, that all have a corresponding word with opposite quantity category.

## 2. Introduction to empirical studies

Besides pointing out stress and quantity as very important pronunciation goals, and letting vowels and consonants be equal as potential carriers of length, a number of details can be found and tested within the BP approach. The studies to be reported in this dissertation all deal with the relation between a phonologic distinction and its phonetic and acoustic correlates, and they are all parts of the scrutiny of the description and priority principles that underlie BP, as elaborated in 1.4 and 1.5.

### 2.1. Segmentation

All studies in this dissertation contain measurements of vowel and consonant durations, and criteria for segmentation is crucial for the reliability of the results. Segmentation has always been performed in Praat (Boersma & Weenink 2001), and waveform, spectrogram and listening have always been utilized in combination. It is admitted that some of the studies in the present thesis are carried out with more eagerness than experience, which sometimes has resulted in the choice of sub-optimal test words, with respect to distinct segment boundaries. In a few cases segmentation has been somewhat problematic, but having utilized the combination of all three analyzing modes mentioned above, it is the present author's firm conviction that the uncertainty in measures have not influenced main trends in the results, or conclusions from them. Examples are finding the boundary between vowel and nasal, and also finding the end of a word-final [a] when the following word starts with ...ja... which often results in a [ɐɐ] without a typical fricative phase for the /j/. In the latter case a boundary is defined half way in the time-dimension between the lowest F2-value for the first [ɐ] and the highest F2-value for the [ɪ] segment. Figure 2.1 shows an oscillogram and a spectrogram of the word *woman*, appearing as test word in section 3.5, a version judged by the present author as showing the most unclear boundaries of the present material. Both the [w] and the [m] have rather indistinct boundaries to the vowel, but listening to a number of sequences ranging over 5-6 segments, moving the listening endpoints 1-2 periods earlier or later, can result in a reliable segmentation, that often coincides with changes in formant character and small differences in amplitude, visible in the oscillogram. The /ʊ/ in *woman* was measured to 97 milliseconds in the result of study 5 (sec-

tion 3.5.3), and measuring with the same methods some months later yielded 99 milliseconds. Other details on segmentation are accounted for in the method chapters of the respective studies.

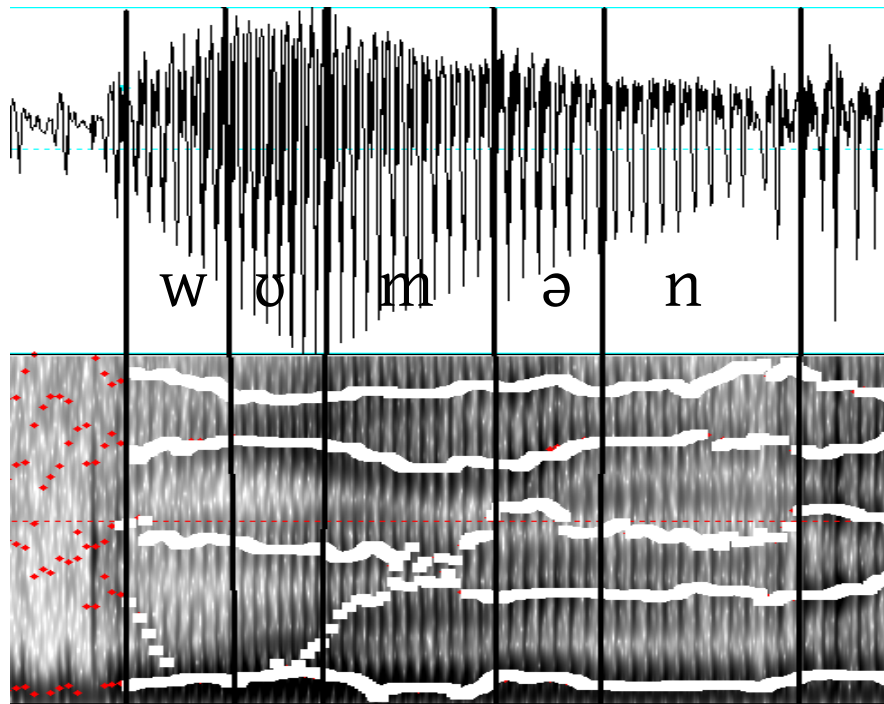


Figure 2.1. Example of segmentation of the word ‘woman’. White curves (enhanced in this picture) are formants showed by the software (Praat).

## 2.2 Studies aiming at the further illumination of the major issues in the concept of Basic Prosody

This dissertation will continue by reporting five studies, aimed to increase our knowledge concerning the timing in stressed syllables in Swedish, and evaluate some of the claims and assumptions in BP. The main questions for the respective studies are:

### 2.2.1. Experiment 1

Does Swedish have 9 vowel phonemes sensitive to duration as a means of perceptually signaling and changing the quantity categories /V:C/ and /VC:/, or is /ʌ/ - as suggested by Hadding-Koch & Abramson (1964) – an exception, only depending on spectral differences between two distinct phonemes; /ʌ/ and /ə/? Manipulations that take the complementary relationship /V:C/ -

/VC:/ into account, i.e. utilizing the possible perceptual effect of postvocalic consonant duration, attempts to shed some light on this issue.

### 2.2.2. Experiment 2

Does the duration of the postvocalic consonant serve as a complementary perceptual cue to the quantity contrast? It seems clear from experiments by Hadding-Koch & Abramson (1964), Traunmüller & Bigestans (1988), Behne et al. (1997, 1998), that vowel duration, and in the case of /u/, vowel spectrum, are the main perceptual cues utilized by native listeners to discriminate between /V:C/ and /VC:/. The first experiment of the present thesis shows that also words containing /u/ can be perceived as “opposite quantity category” by means of mere durational changes, in the mentioned study, see 2.2.1. This experiment uses two series of stimuli; one with manipulated vowel durations only, and one with altered complementary consonant durations added to the series of altered vowel durations. A comparison of the results from the both series will show whether there is a contribution from the consonant duration to listeners’ quantity categorization.

### 2.2.3. Production study 3

Can an L2 user signal the Swedish quantity contrast with other acoustic correlates, or other weights of existing correlates than those typical to the native Swedish speaker? The existing correlates are assumed to be duration of vowel and postvocalic consonant, and vowel spectrum. Can unambiguous quantity categories be signaled by means of i) vowel duration only, ii) vowel spectrum only, or iii) consonant duration only? Will a number of L2-realizations of the Swedish quantity contrast support or contradict the importance of consonant duration (1.4.3), or the notion of pedagogical elasticity, mentioned in 1.5.3?

### 2.2.4. Production study 4

Can an L2 user signal the Swedish word stress contrast with other acoustic correlates, or other weights of existing correlates than the native Swedish speaker? The existing correlates are assumed to be syllable duration, tonal gestures, intensity, spectral tilt and vowel tenseness. The study puts focus on temporal properties and the possible role of the VC-sequence as carrier of the increased syllable duration, claimed to be the most reliable correlate to stress in Swedish (Fant & Kruckenberg 1994). The role of the postvocalic consonant, as well as aspects of elasticity, will be taken into account.



### 2.2.5. Production study 5

Is the complementary postvocalic consonant length such a persistent feature of Swedish prosody, that it is realized also when native Swedes speak L2's that do not utilize or otherwise show this complementary temporal organization? In this study, some native Swedes pronounce words in English and German. The words are assumed to be perceived by the Swedish speakers as having "short vowel quantity", and the question is: Will the Swedish speakers lengthen the postvocalic consonant when they pronounce the test words in English and German?

### 3. Empirical studies

#### 3.1 Duration versus spectrum: A perception test with manipulated durations.

##### 3.1.1. Introduction

Out of the need to simplify linguistic description for pedagogic purposes, a basic prosodic strategy (BP) for teaching Swedish L2 pronunciation has developed over the last 2-3 decades. The development of the BP-strategy is described in detail in 1.3 and 1.4. Swedish is a language with three distinctive prosodic contrasts: word stress, e.g. *planet* [pló:nət] ‘the plane’ – *pla’net* [pléné:t] ‘planet’, and distinctive length i.e. quantity, e.g. *glas* [gló:s] ‘glass’ – *glass* [glás:] ‘ice cream’, in addition to a tonal word accent; *anden* ‘the duck’ – *ànden* ‘the spirit’. All three play a role in learning (and teaching) Swedish as an L2, although word stress and quantity are in this thesis considered as contributing more to a listener friendly pronunciation than the tonal word accent. See discussion in 1.5. The experiment presented here deals with the question of whether duration or spectrum is the main perceptual cue to the Swedish quantity distinction, and also whether manipulation of vowel and consonant duration in combination, can make a native Swedish listener perceive a word containing originally long vowel as the corresponding word with short vowel (and vice versa), only by means of temporal change, preserving the spectral properties. The classical experiment in this area was carried out by Hadding-Koch & Abramson (1964). They changed the duration of vowels and stop consonants by means of cutting and splicing magnetic tape, and they manipulated vowels by cutting out pieces of tape from the central part of the vowel, where formant frequencies were relatively steady. The voiceless stop /t/ had its silent occlusion phase changed between long and short allophone, rendering the word *stöta* [stø:ta] ‘push’ the same /t/-duration as the word *stötta* [støet:a] ‘prop up’, and vice versa. Reducing vowel duration resulted in native Swedish listeners perceiving original /V:C/-words as /VC:/-words, except for words containing /u/. Changing consonant duration did not make the listeners perceive the corresponding word with non-original quantity category. They concluded that a) consonant dura-

tion is not a primary cue to the quantity contrast, and b) that the /ʉ/ phoneme is not one phoneme with a long and a short allophone, rather was it considered as two separate phonemes /ʉ/ and /ø/, kept apart by spectral properties. The present study is carried out with partly the same setup as Hadding-Koch & Abramson (1964), but the subjects are from middle and North Sweden (compared to Hadding-Koch & Abramson's South Swedish subjects), and the temporal manipulations involve both vowel and consonant duration according to the complementary pattern, i.e. an increase in vowel duration is accompanied by a decrease in consonant duration and vice versa. The manipulations in the present experiment are extended beyond typical values for /V:C/-words and /VC:/-words respectively, to find out whether e.g. the vowel /ʉ/ is sensitive at all to temporal manipulation. Another aim is to scrutinize the suggestion of Hadding-Koch & Abramson (1964), that Swedish has "...a subset of eight vowel phonemes that combine with a phoneme of length to yield long counterparts and another subset of two vowel phonemes, /ʉ/ and /ø/, distinguished by quality."

Elert (1964) showed that the quantity distinction has temporal correlates, and also, that the quantity is realized as a complementary durational relation between vowel and subsequent consonant in stressed syllables; /V:C/ - /VC:/. The BP approach focuses on stress and quantity, and urges the learner to lengthen the correct speech sound, thus regarding 'length' as both a phonological property and physical realization. The teacher and the learner are not expected to distinguish phonology and correlates in the educational situation, and BP wants to take advantage of the fact that phonological length and duration can be confused without causing any problems. If the learners of Swedish as an L2 implement the length in the phonologically long sounds as increased duration, it is assumed that a native Swedish listener easier perceives both the quantity category and the word stress category, compared to a strategy that would exhort learners to use vowel spectrum as main realization of quantity and tonal gesture as main realization of stress. See 1.5.1-1.5.3 for a review of empirical findings and a discussion on this topic.

The phonological distinction /V:C/ - /VC:/ is also accompanied by a spectral difference in the vowel, which is greater for some vowels and smaller for others (e.g. Elert 1970). These spectral differences are more or less ignored in BP, and this is accounted for in 1.5.2. Duration and spectrum also interacts differently in different regional varieties of Swedish, and e.g. Reuter (1971) finds that Finland-Swedish (Helsinki variety) has smaller spectral differences between long and short vowel allophones than central standard Swedish, which is noticeable for the vowel phonemes /a/ and /ʉ/, both of which differ greatly between long and short allophone in most regional varieties of "Sweden-Swedish". Reuter (1982) found furthermore that Finland-Swedish long/short vowel allophones differ more in duration than is the case for central standard Swedish.

The pedagogical BP approach, or at least how it is commonly implemented suggests that temporal correlates to the Swedish quantity are more important cues than spectral. “More important” should here be interpreted as more important as perceptual cues and hence more important to teach/learn. This is an assumption based on experience, in combination with the regional and between-vowel differences, which imply that duration as correlate to the Swedish quantity distinction, is a factor that unites regional varieties (dialects), as well as vowel phonemes, and spectral differences seem to be more dialect specific and vowel specific.

A study by Behne et al. (1997) suggests that duration is a stronger cue to discrimination for /i/, /o/ and /a/, but that the interaction between duration and formant spectrum for /a/ is complicated.

The main purpose of the present study is to test whether the temporal correlates of the Swedish /V:C/ - /VC:/-distinction is a stronger cue to Swedish quantity than the spectral correlates, and to compare a method using complementary duration to studies using manipulation of vowel duration only, e.g. Behne et al. (1997) and Hadding-Koch & Abramson (1964).

### 3.1.2. Method

#### 3.1.2.1 Stimuli

12 monosyllabic Swedish words, providing 6 minimal word pairs, each pair having identical phoneme sequences and differing only in quantity, were read by two native Swedish (Stockholm variety) speakers, one male and one female. The words were: mat-matt [mɔ:t-mat:] ‘food-faint adj.’, vit-vitt [vi:t-vit:] ‘white’, two gr. genders, hut-hutt [hʉ:t-het:] ‘manners-snifter’, våt-vått [vɔ:t-vɔt:] ‘wet, two gr. genders’, tät-tätt [tɛ:t-tɛt:] ‘tight/close’, two gr. genders, and söt-sött [sø:t-søet:] ‘sweet’ two gr. genders. This common way of using different IPA-characters for the long-short allophones of the Swedish vowels reflects the spectral difference that is noticeable for most of the vowel phonemes.

The words were read in alphabetic order from a list and recorded in a studio at the Phonetics department of the Department of Linguistics at the University of Stockholm. Based on duration values from the readings of the two speakers, a duration of 600 ms was chosen as total duration of the VC-sequence for all stimuli in the experiment. The recording of the female speaker was synthesized and manipulated in Praat (Boersma & Weenink 2001). Praat manipulations by the PSOLA-method allows for reducing and increasing durations evenly in a chosen section. Each of the 12 words was given V and C durations shown in table 3.1.1, actually just moving the boundary between V and C, thus preserving the total VC-duration of 600 milliseconds. Segmentation was made by means of spectrogram, oscillogram and listening (cf. 2.1.). When the words had voiceless initial consonant, the

vowel duration was defined as the voiced part of the word, and when the words (*mat*, *matt*, *våt*, *vått*) had voiced initial consonant, the boundary between consonant and vowel was not problematic. The final t-segment was defined as the beginning of the voiceless part after the vowel and included the burst. All long vowel phonemes are in central standard Swedish more or less diphthongized, and in the present material, the vowels /i:/ and /ɛ:/ are the phonemes with the most characteristic closing phases towards the end of the vowel; palatal [ij] for /i:/ and labial [ɥβ] for /ɛ:/. Studies of spectrograms and listening control were carried out after the manipulations, to certify that these spectral characteristics of the long vowel allophones were preserved also in the shortest versions used in the present study.

Table 3.1.1. Vowel durations, consonant durations (milliseconds) and V/C-ratios for all test words.

V/C ratio in the Diagonal		Consonant duration						
		250	300	350	400	450	500	550
Vowel duration	350	1,4						
	300		1,0					
	250			0,7				
	200				0,5			
	150					0,3		
	100						0,2	
	50							0,1

### 3.1.2.2 Presentation

7 versions of each word made a total of 84 stimuli, which were presented in random order to 20 native adult Swedish listeners. Presentation was made with headphones from computer or CD-player. The listeners heard each stimuli once, and responded on an answering sheet whether they perceived it as a /V:C/-word or as a /VC:/-word, by marking one of two words, e.g. *mat* or *matt*, with forced choice. After the categorization test, the listeners judged the naturalness of the pronunciation for each stimulus by means of the VAS method (Visual Analogue Scale, Huskisson 1974, Wewers and Lowe 1990), with printed numbers from 1 “totally unnatural pronunciation” to 10 “totally natural pronunciation”.

### 3.1.3. Result

#### 3.1.3.1 Original V/C-ratios

Table 3.1.2 shows ranges and mean values of the V/C-ratio for the original, intact recordings of the test words. In these recordings there seems to be a “vacuum” of V/C-ratios between 0.40 and 0.50, which could be interpreted

as a “security distance” for the discrimination. The total number of six recordings (one male and one female speaker reading at intended slow, medium and fast rate) show in fact very few observations between 0.30 and 0.70.

Table 3.1.2. Range and mean values for V/C-ratios in original recordings.

	V/C-ratios for /V:C/ words		V/C-ratios for /VC:/ words	
Mat – matt	0.67	– 0.97	0.20	– 0.38
Vit – vitt	0.51	– 0.89	0.23	– 0.25
Hut – hutt	0.66	– 0.83	0.19	– 0.32
Våt – vått	0.63	– 1.10	0.15	– 0.37
Tät – tätt	0.68	– 1.02	0.25	– 0.32
Söt – sött	0.72	– 1.06	0.23	– 0.37
Total and mean	0.51	– 1.10    m=0.79	0.15	– 0.38    m=0.27

### 3.1.3.2 Categorization

The result of the categorization test is captured in figure 3.1.1, and shows that all included vowel phonemes are influenced by the changes of vowel and consonant durations, expressed as V/C-ratios, meaning that most, if not all, native Swedish listeners changed their categorization from /V:C/-word to /VC:/-word and vice versa, as a result of temporal changes. From figure 3.1.1 it is clear that *hut-hutt* needed great temporal changes in order to be perceived as ‘non-original-category’. For original *hutt* (/VC:/), just over 50% of the listeners categorized stimuli as *hut* (/V:C/), when the VC-ratio was 1.4, i.e. nearly a doubling of the original V/C-ratio of 0.66 – 0.83 for *hut*. The word pair *tät-tätt* constitutes the other extreme, by showing a steep change in categorization around V/C 0.4, with roughly the same crossover value for both /V:C/-word and /VC:/-word as starting points.

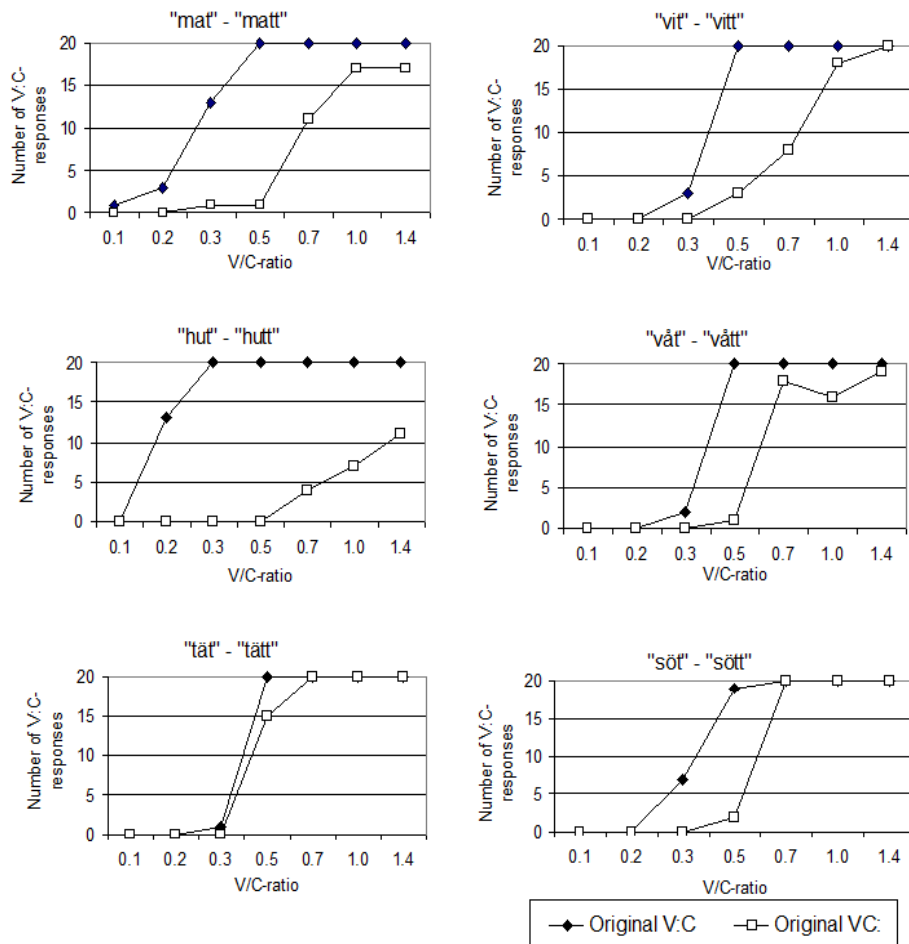


Figure 3.1.1. Number of listeners perceiving stimuli as /V:C/-word as a function of V/C-ratio. A solid line with filled squares always represents an original /V:C/-word and a solid line with open squares always represents an original /VC/-word.

### 3.1.3.3 Naturalness

Judgments of naturalness are presented in figure 3.1.2 and 3.1.3. As could be expected, naturalness judgments decrease as the V/C-ratio differs more from the typical combination of temporal value and spectral character of the intact production of /V:C/ and /VC/-word respectively. This decrease is more obvious for original /VC/-words. The vowel phonemes / $\epsilon$ / and / $\emptyset$ /, traditionally known as having minimal spectral difference between long and short allophones are judged as highly natural both when perceived as /V:C/ [t $\epsilon$ :t], [s $\emptyset$ :t] and when perceived as /VC:/ [t $\epsilon$ t:], [s $\emptyset$ t:], with the high naturalness judgments more evenly distributed when original word is /V:C/-category. This tendency is true for phonemes /i/, /o/, / $\epsilon$ / and / $\emptyset$ /, indicating that their original /VC/-spectrum is more marked than the /V:C/-spectrum. Original

“hutt” [høt:] is judged as natural only when having V/C-ratios 0.1 to 0.5, i.e. V/C-ratios similar to original version read by the Swedish speakers.

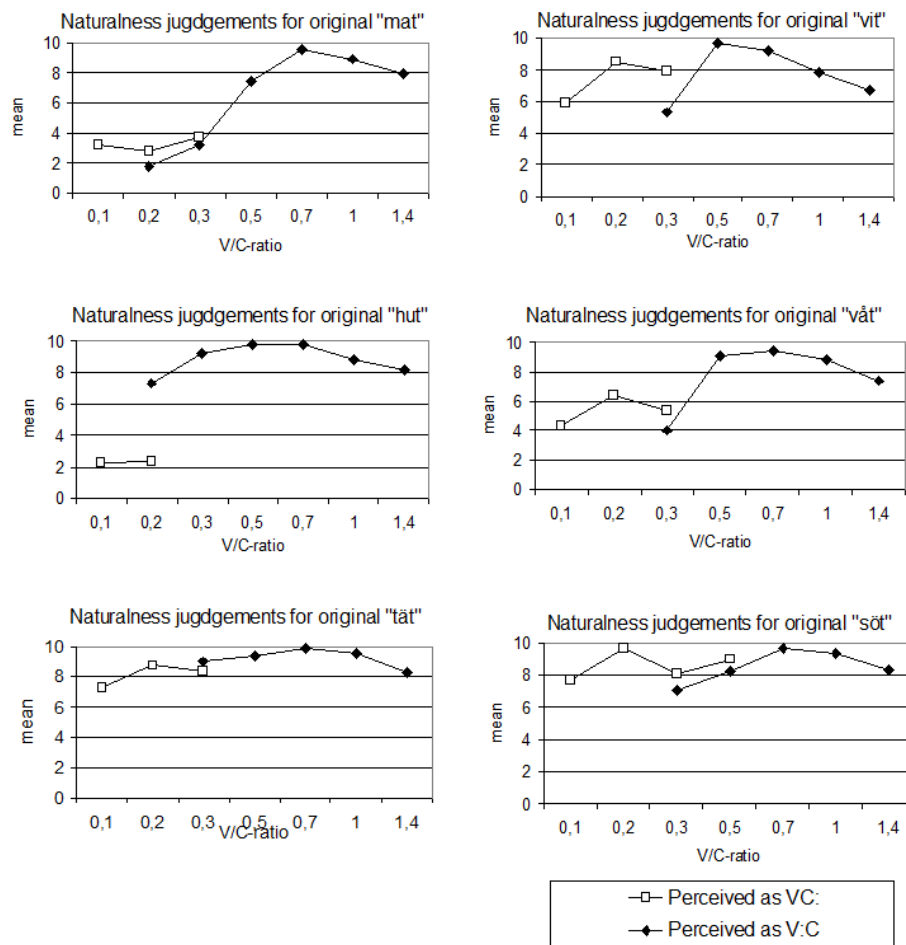


Figure 3.1.2. Mean values of naturalness judgements for original /V:C/ words at different V/C-ratios and perceived as either /V:C/ or /VC:/



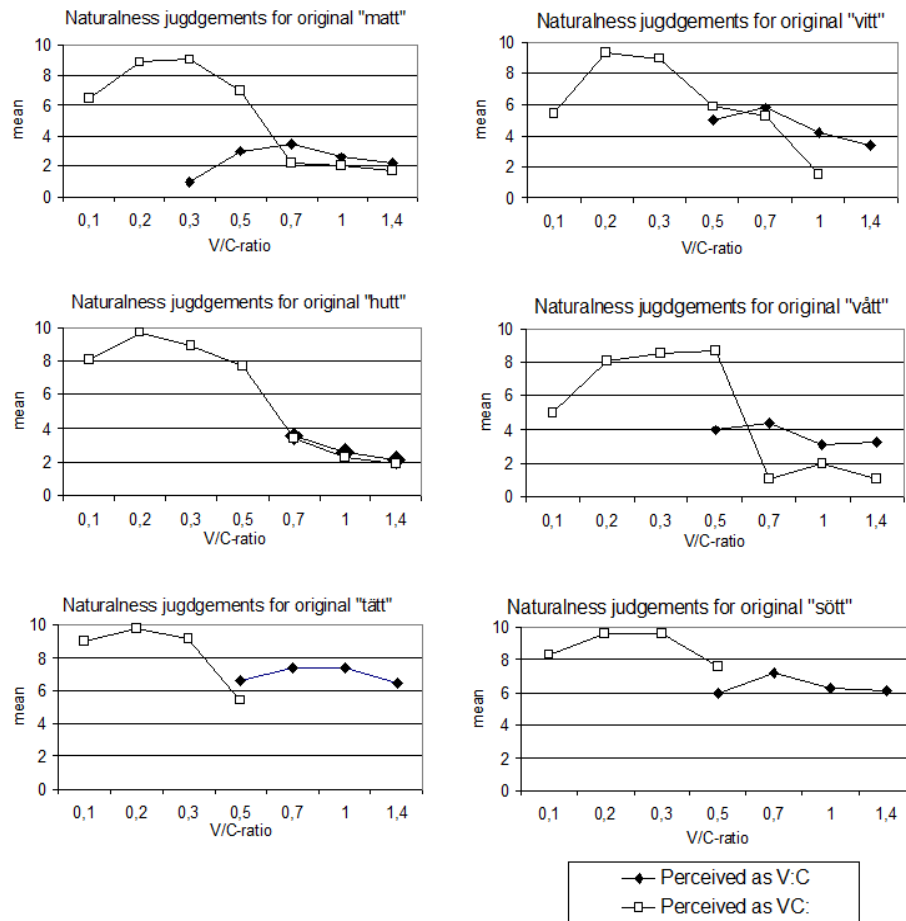


Figure 3.1.3. Mean values of naturalness judgements for original /VC:/ words at different V/C-ratios and perceived as either /V:C/ or /VC:/

### 3.1.3.4 Duration versus spectrum

So far we have stated that native listeners can perceive an original /V:C/-word as a /VC/-word, and vice versa, with the aid of temporal change only, and also that this is more prevalent for some vowel phonemes than for others. This material does not allow for a reversed test, namely to keep V/C-ratio constant and gradually change the vowel spectrum. We can, however, combine a V/C-ratio typical for one quantity type, with the spectrum typical for the other quantity type (namely the original spectrum for either of the two quantity types). Will the listeners assign the word the quantity category in accordance with its spectral or its temporal character?

The typical ratios for /V:C/-words are best represented by ratios 0.7 and 1.0 in the original recordings, and the typical ratios for /VC/-word are best represented by ratios 0.2 and 0.3. These V/C-ratios have also received the

highest ratings for naturalness. Let us compare the number of /V:C/-judgements (averaged over the two mentioned values) for each vowel phoneme at the typical V/C-ratio values for long and short vowel respectively, and see how many listeners perceive stimuli as non-original quantity type, when spectral properties are shifted at constant V/C-ratio.

Figure 3.1.4 shows how changes in temporal and spectral properties respectively affect the perception of the listeners. The average responses over all six vowel phonemes show that temporal change has caused more shifts in perception than spectral change. The /ʊ/ phoneme is the only one that is more affected by spectral than temporal change, and /ɛ/ and /ø/ are in these figures unaffected by spectral change, but have caused all listeners to shift their perception from original to non original quantity type as a result of temporal change.

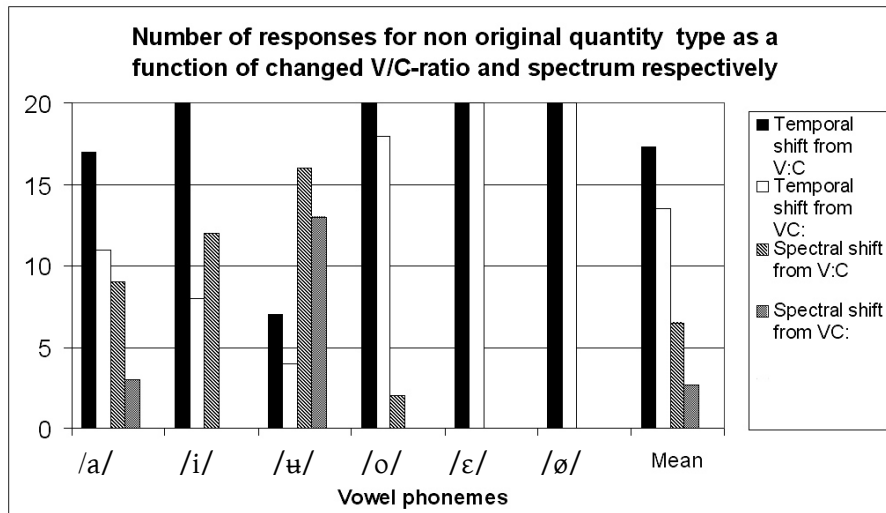


Figure 3.1.4. Number of listeners changing their judgements to non-original quantity type as a function of temporal and spectral manipulation.

### 3.1.4. Discussion

The results indicate that all of the included vowel phonemes use both duration and spectrum as cues for the quantity distinction, but in different proportions for various vowel phonemes. The different phonemes show varying sensitivity to the temporal change. The vowel that seems most apt to be perceived as the non-original quantity type due to mere temporal change is /ɛ/, and most resistant to temporal change is /ʊ/. Most of the shifting over from original to non-original quantity type takes place between V/C-ratios 0.3 and 0.7, which corresponds rather well to the “vacuum” observed in the original recordings of native speech. We can note that short [ɪ] and short [ə] need

V/C-ratios greater than the typical ones, in order to be perceived as long, and that long /v:/ and long /ʉ:/ need V/C-ratios smaller than the typical ones, in order to be perceived as short.

Since the various phonemes differ to a large extent in their duration/spectrum dependence, and since all of the vowel phonemes are influenced by temporal change, the results do not support the suggestion of Hadding-Koch & Abramson (1964), that the /ʉ/ phoneme would hold an exceptional position in the Swedish vowel inventory, but it puts /ʉ/ at one end of a continuum of “duration/spectrum dependence” and the /ɛ/ phoneme at the other end, /ʉ/ being more dependent on spectrum than on V/C-ratio, and /ɛ/ being more dependent on V/C-ratio than on spectrum. The fact that Hadding-Koch & Abramson (1964) used south Swedish speakers and listeners can also explain some of the differences in results and conclusions between their study and the present one. If we consider the outcome for original *hut* and *hutt* respectively (figure 3.1.1), we see that lengthening of short vowel does not result in all listeners changing their categorization, but going from long to short vowel they did, albeit at a very small V/C-ratio. This might lead to the suspicion that the vowel in original *hut*, at a V/C-ratio of 0.1 (corresponding to a duration of 50 ms) was not long enough to allow the listeners to perceive the spectral qualities of the vowel segment, while the lengthened vowel in original *hutt* did. The sharp decrease in perceived naturalness as listeners shifted from original *hut* to *hutt* implies that the listeners did perceive the vowel quality well enough.

The six studied phonemes could be lined up according to their respective dependence on V/C-ratio or spectrum, using the diagrams in figure 3.1.4, with number of listeners categorizing according to temporal change / number of listeners categorizing according to spectral change:

/ɛ/ 40/0, /ø/ 40/0 >> /o/ 38/2 >> /i/ 28/12, /a/ 28/12 >> /ʉ/ 11/29

The results suggest that both temporal and spectral properties play an important role for Swedish V:C/VC-categorization, but taken over the vowel inventory in this study, the temporal organization seems to be a stronger cue to the listeners' perception than does the spectrum. We see also, as shown in table 1.4, and discussed in 1.5.2, that the speakers make similar temporal differences for the /ʉ/-phoneme, as for the other vowel phonemes, in spite of the little distinctive use native listeners seem to make of duration with respect to this phoneme. This can indicate that not only /V:C/ – /VC:/ discrimination must be maintained, but also durational differences between stressed and unstressed syllables (cf. Behne et al. 1998, Fant & Kruckenberg 1994). The increased duration in stressed syllables seems to have at least two roles to play; one to distinguish stressed syllables from unstressed ones, and one to distinguish the two quantity categories. The conclusions for pedagogical purposes is that the temporal organization is of superior importance

because of its double function, and that the spectral differences are also important, but play a role only in quantity distinction, and not substantially for all vowel phonemes. The BP-priority of length over spectrum, also leads to one single distinction for all vowel phonemes, namely ‘very long’ and ‘very short’, compared to a length difference plus roughly 18 different spectral qualities, if one gives the same priority to both length and timbre (practical terminology in language instruction), or duration and spectrum (phonetic terminology).

Compared to the results of Hadding-Koch & Abramson (1964), it is clear that the present study managed to make listeners change their perception of quantity category by means of mere temporal change, for all included vowel phonemes, including /u/, and since at least four conditions are changed since the former study, one cannot say which factor made it possible to show the duration-sensitiveness in /u/. First, the present study used manipulations of vowels and consonants in combination, utilizing the complementary pattern, allowing for a possible impact of the consonant duration as a complementary perceptual cue for the listeners. Second, the manipulations were gradually producing durations and V/C-ratios beyond the values of the ‘target values’ i.e. the durations typical to the other counterpart of the minimal quantity word pair. Third, both the speakers and the listeners of the present study were from various parts of Sweden, except the province of Skåne, where the Hadding-Koch & Abramson (1964) study was conducted, and known for a dialect with many deviations from central standard Swedish, not least vowel qualities and temporal organization (cf. 1.4.3, 1.5.2. and Gårding et al. 1974). Fourth, the test words in the present study were pronounced in isolation, whereas Hadding-Koch & Abramson (1964) used a carrier phrase with the test word in non-final position, preventing the word final consonant to be affected by final lengthening, a phenomenon that may have affected the outcomes of the present study.

Considering the present results and Reuters’ (1971) findings, it would be recommended, that more similar experiments be carried out with various Swedish regional varieties, to extract what is general Swedish quantity, and what is regional or local Swedish quantity.

## 3.2. The postvocalic consonant as a complementary cue to the perception of quantity in Swedish – a revisit

### 3.2.1 Introduction

The BP-approach can in the educational implementation be reduced to a short recommendation: “lengthen the proper speech sound”, thus aiming at enhancing the word stress as well as the quantity contrast, both of which

depend mainly on duration as perceptual cue for the listener (cf. the discussion in 1.5.2.). Measuring of Swedish syllable duration has shown that stressed syllables are 50-100% longer than unstressed syllables (e.g. Strangert 1985, Thorén 1982). If a stressed syllable containing a short vowel is going to be lengthened, an increased post-vocalic consonant duration is one way of maintaining the proper duration of the stressed syllable. The general consensus, based on findings from e.g. Hadding-Koch & Abramson (1964), Behne et al. (1997, 1998), is that the vowel duration is the main perceptual cue to the quantity contrast, but at the same time we know that stressed syllables have longer duration than unstressed ones, and the duration must be attached to some part of the speech. Fant et al. (1991b) showed that stressed syllables of /V:C/-type are lengthened mainly by means of vowel duration, while stressed /VC:/-syllables are lengthened by means of the complementary postvocalic consonant duration. The question is, does the duration of the postvocalic consonant have any role as perceptual cue to the quantity distinction, or is it primary a buffer zone, to render the stressed syllable its proper duration? This is discussed in 1.4.3, and 1.5.2. From the pedagogical perspective of BP, it seems reasonable to include the complementary consonant length in order to maintain the natural timing of Swedish, irrespective of its possible role as perceptual cue to phonological contrasts. Thorén (2001) showed that digitally increased duration in phonologically long segments in Swedish with a foreign accent tended to be judged as improved Swedish pronunciation by native Swedish listeners. The study showed similar effect for lengthening of both vowel and consonant duration, and greatest effect in sentences containing both vowels and consonants that had been lengthened. The non-native speaker who had produced the original material in the study was Polish, and Polish is a language without phonological quantity. Although the total result was significant, it should be observed that the native Swedish listeners were not always able to distinguish intact versions from manipulated versions. The present experiment aims at testing the role of the postvocalic consonant as a perceptual cue to the Swedish quantity contrast. If it will be shown that the postvocalic consonant plays a role also as cue to the quantity distinction, it would provide additional arguments in favor of its pedagogical value.

The first study – known to the present author – that involves testing duration of the postvocalic consonant as a cue to the quantity contrast, is Hadding-Koch & Abramson (1964), described in 3.1. They lengthened the occlusion of the postvocalic /t/ in *stöta* [stø:tə] ‘push’ and shortened the occlusion of the postvocalic /t/ in *stötta* [støt:ə] ‘prop up’ and tested whether native Swedish listeners would perceive original *stöta* as *stötta* as a result of stop closure lengthening, and vice versa. The Swedish listeners did not change their categorization of the words, as a result of mere changing of stop closure duration. Behne et al. (1998) attempted to minimize the effect of vowel duration and vowel spectrum, when testing the perceptual influence of

the postvocalic consonant duration. In a L2-pedagogical perspective, this can be considered highly relevant, since L2-learners of Swedish – in the present author’s long experience – seldom master the combination of vowel quality and vowel duration, that are typical to the Swedish quantity categories, thus producing a language with vague signaling of quantity. Behne et al. (1998) produced stimuli that were synthesized from natural recordings, resulting in vowel duration and formant spectra that were halfway between the typical values for /V:C/-words and /VC:/-words, thus attempting to minimize the impact from the vowel as a cue to categorization. The vowel phonemes were /a/ /o/ /i/. In one series the vowel was followed by /t/, and in the other by /d/. No immediate effect of consonant duration could be seen, and they conclude: “...listeners do not use the duration of postvocalic consonants to identify vowel quantity in Swedish, even in the absence of durational or spectral information from the vowel...” (Behne et al. 1998: 97).

The present study is an attempt to further refine the testing of the possible influence of the postvocalic consonant to the quantity distinction, by using various vowel durations, from typical (i.e. intact) values for /V:C/-words, and stepwise to typical values for /VC:/-words. As has been shown in section 3.1, most Swedish vowels are sensitive to duration change, with respect to quantity categorization by native Swedish listeners, and in a first attempt Thorén (2004) used the recordings from Thorén (section 3.1.), with isolated words read from a list. The series of stimuli that were manipulated as shown in table 3.1.1 were supplemented with a series that only had their vowel durations changed, with postvocalic stop durations intact. This way the effect of stepwise change of vowel duration could be compared to the effect of stepwise change of vowel duration in combination with complementary stepwise change of postvocalic consonant duration. The vowels tested in Thorén (2004) were /a/ and /ʌ/ representing vowels with great spectral differences between long and short allophone, and /ɛ/ and /ø/, representing vowels with minimal spectral differences between long and short allophone. If there would be any difference in categorization made by native Swedish listeners, between two versions with the same vowel duration, this difference must be ascribed the influence of the postvocalic consonant, being the only factor that is not identical in the both versions. The result of Thorén (2004) is that the duration of postvocalic consonant duration had an influence on the listeners’ categorization on the case of /a/ and /ʌ/, but not for /ɛ/ and /ø/. The greatest difference between the two manipulation series is at a vowel duration typical to non-original quantity category and not at the vowel duration closest to the expected mid-point between typical values for /V:C/ and /VC:/-words. It is also admitted that Thorén (2004) missed to include this mid-point in the setup, and that is one of the reasons to conduct this revisit. Another reason is to put the test words in a carrier phrase, since it can be assumed that a native listener uses clues outside the VC-domain to perceive quantity category (cf. Diehl & Kluender 1987: 227 and the discussion in

1.6.). In a study by Traunmüller & Bigestans (1988) it was shown that vowel duration related to the duration of the entire utterance (in this case, carrier phrase), was the measure that best matched the listeners' categorization into /V:C/ and /VC:/-words. To add a carrier phrase could be expected to make the listeners perceive quantity categories in a more natural way, although the only differences that would be studied is that of postvocalic consonant duration.

The present study compares two series of vowel duration manipulations; one with changing of vowel durations only, and one with vowel and consonant duration change in combination, in accordance with the complementary VC-relation in Swedish. This method could evaluate the consonant duration as a possible complementary cue to the Swedish quantity distinction.

- Hypothesis 1: Complementary vowel + consonant duration change helps the listener perceive the “non-original” quantity category with less vowel duration change, than would be the result of change of vowel duration only.
- Hypothesis 2: Test words with complementary duration – /V:C/ or /VC:/ – will be judged as more natural sounding than words with “correct” vowel duration and “wrong” consonant duration i.e. short /VC/ or over-long /V:C:/.

### 3.2.2 Method

#### 3.2.2.1. Stimuli

The test words in the present study are *mäta* [mè:tə] ‘to measure’, *mätta* [mèt:ə] ‘to satisfy’ *skuta* [skù:tə] ‘boat’, *skutta* [skèt:ə] ‘to scamper’. These words provide two minimal pairs with respect to phonological quantity. One pair contains the vowel phoneme /ɛ/, and the other pair contains the vowel phoneme /ʉ/. The words were recorded in a sound treated room in the present authors home, using a Røde NT3 condenser microphone and a Sony MZ-N710 mini-disc player. The speaker was a Swedish male, speaking central standard Swedish (Stockholm variety). The test words were pronounced within a carrier phrase: *Det var ..... jag menade* ‘It was ..... that I meant’ Vowel and consonant durations in the test words were manipulated in Praat (Boersma & Weenink 2001). All stimuli were given stepwise vowel duration change. Half of the stimuli kept a constant consonant duration, identical with the original quantity category, and the other half were given stepwise consonant duration changes, based on original values for non-original quantity category. The manipulated durations are shown in table 3.2.1.

Table 3.2.1. Vowel- and consonant (occlusion) durations for manipulated stimuli in the present study. Shaded parts represent original durations for non-original quantity type.

<b>Changing of vowel duration only (ms)</b>							
Original	V	188	168	148	128	108	88
[mè:tɐ]	C	153	153	153	153	153	153
Original	V	136	156	176	196	216	236
[mèt:ɐ]	C	334	334	334	334	334	334
Original	V	141	121	101	81	61	41
[skù:tɐ]	C	166	166	166	166	166	166
Original	V	166	186	206	226	246	266
[skèt:ɐ]	C	312	312	312	312	312	312
<b>Changing of V and C duration (ms)</b>							
Original	V	188	168	148	128	108	88
[mè:tɐ]	C	234	254	274	294	314	334
Original	V	136	156	176	196	216	236
[mèt:ɐ]	C	253	233	213	193	173	153
Original	V	141	121	101	81	61	41
[skù:tɐ]	C	232	252	272	292	312	332
Original	V	166	186	206	226	246	266
[skèt:ɐ]	C	246	226	206	186	166	146

### 3.2.2.2. Listeners

30 native speakers of Swedish listened to the 48 stimulus words, marking whether they perceived them as /V:C/ or /VC:/. The listeners were between 23 and 60 years of age, and had different regional varieties of Swedish as their L1. None of them had any hearing deficiencies that affected their perception of normal speech.

### 3.2.2.3. Presentation

The 48 stimuli were presented in random order, in the carrier phrase, preceded by the reading of stimulus number. The test was presented from CD-player via headphones. The listener was first allowed to hear 2-3 stimuli while adjusting the sound level. The response was marked on an answering sheet, presenting the stimulus number and the pair of words providing the two choices. The listener had to make a forced choice between one of the two possibilities. Naturalness rating was done in direct connection to each categorizing task. After hearing the test word a second time, the listener marked a figure (1-10) on a VAS-scale (see 3.1.2.), where 1 represented “totally unnatural or unlikely pronunciation for a native speaker of Swedish” and 10 “totally natural pronunciation for a native speaker of Swedish, regardless of regional variety”.



### 3.2.3. Result

#### 3.2.3.1. Categorization

In both the vowel lengthening series and the vowel shortening series, the complementary consonant manipulation seems to have an influence on the listeners perception of /V:C/ or /VC:/ (figure 3.2.1). Listeners start to perceive stimuli as non-original quantity category at lower degree of vowel duration change, when the post-vocalic consonant duration follows the complementary pattern. For /ʉ/, the complementary manipulation seems to make less difference compared to /ɛ/, both when going from /V:C/ words to /VC:/ and vice versa.

The over all effect of duration change is greater for /ɛ/ than for /ʉ/, which is expected, because of the greater difference in formant spectrum between long and short allophone of /ʉ/. Recall the earlier findings of the previous experiment, in section 3.1 as shown in figure 3.1.1. More unexpected is the greater effect of postvocalic consonant duration on /ɛ/ compared to /ʉ/ in the present study, which is quite the opposite to the result in Thorén (2004), where the effect was greatest for /ʉ/, and practically none for /ɛ/.

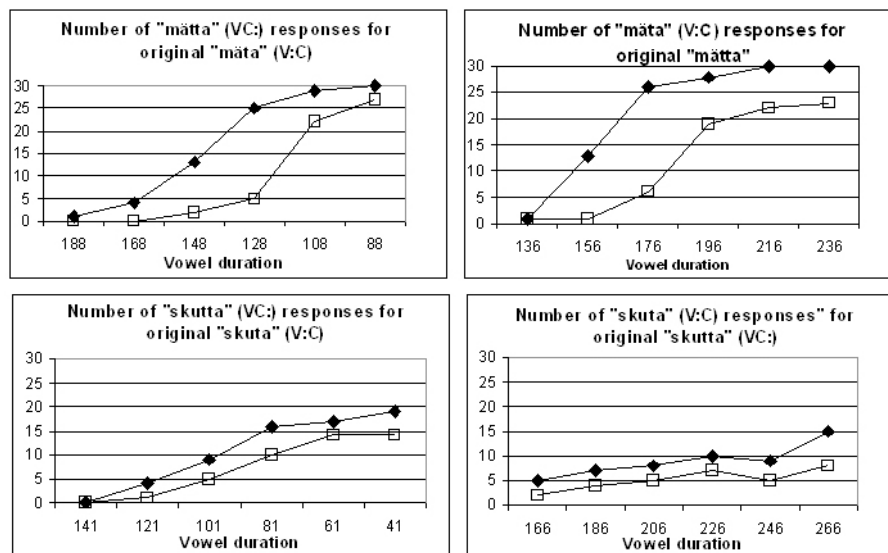


Figure 3.2.1. Number of /VC:/-responses for each value of vowel duration in original /V:C/-words, and number of /V:C/-responses for each value of vowel duration in original /VC:/-words. Filled diamonds represents manipulations of both vowel and consonant durations and open squares represents manipulations of vowel duration only.

### 3.2.3.2. Naturalness rating

“Correct” consonant duration gave mostly higher naturalness ratings in the two /ε/ series, but had a vague effect in the /ʉ/ series, as shown in figure 3.2.2. There was a slight positive effect when going from original *skutta* [skøt:a] to *skuta* [skø:ta] (lower right panel), and a small but consistent negative effect when going in the other direction (lower left panel). The observed effect on naturalness from post-vocalic consonant duration in both series containing the /ʉ/ phoneme has low significance, due to the smaller number of “non-original quantity type” responses.

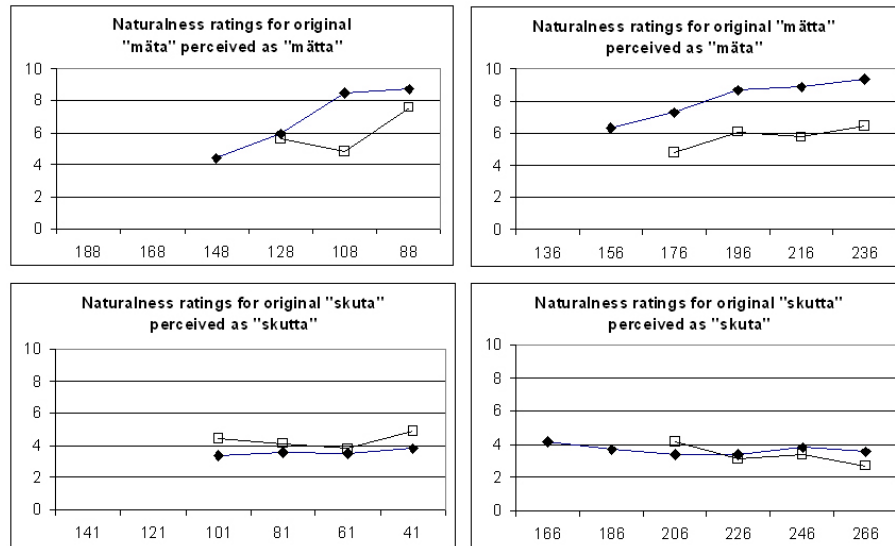


Figure 3.2.2. Naturalness ratings for all versions of the test words. The abscissa shows vowel durations. Open squares represent manipulated vowel duration with intact consonant duration. Filled diamonds represent manipulated vowel durations in combination with complementary manipulations of consonant durations.

### 3.2.4 Conclusion and discussion

The result shows that the duration of the post-vocalic consonant is more than a means to assign the proper length to stressed syllables. It does obviously play a distinctive role for the perception of quantity category in the present material. Since the vowels in this study represent the maximal (/ʉ/) and the minimal (/ε/) spectral differences between long and short vowel allophones in the Swedish vowel inventory (see 3.1.4), the result indicates that the duration of the post-vocalic consonant functions as a general complementary cue to the perception of quantity category in Swedish.

The present experiment and Thorén (2004) both showed some effect of consonant duration on the listeners’ perception of quantity category, and in

both studies words containing /ʉ/ needed more change of vowel duration – with and without accompanying change of consonant duration – than words containing /ɛ/, to be perceived as non-original quantity category. The two studies differ however with respect to the effect of the postvocalic consonant duration on /ɛ/ and /ʉ/ respectively. In both studies it seems clear that vowel duration has a greater impact on the listeners' perception, than consonant duration, which is expected from previous studies. The principal differences between the two studies are a) Thorén (2004) used monosyllabic words, and the present study uses disyllabic words. b) Thorén (2004) studied words read in isolation, while the present study examines word in a carrier phrase containing five words, with the test word in the middle, i.e. as the third word. If the discrepancy between the studies could be explained by total irregularity and mere chance, the graphs for the two manipulation series, as shown in figure 3.2.1, would have crossed over each other and not been held apart in the consistent way that is the case. The fact that the manipulation series in the present study include the whole range where the listeners started to change their perception of quantity category, and Thorén (2004) did not, can possibly account for the discrepancy. In the latter study, almost all listeners – in the case of /ɛ/ and /ø/ - responded non-original quantity category for all examined vowel and consonant durations but one, and in that case it is obvious that versions with adjusted consonant duration received more responses for non-original quantity category, than did the version with same vowel duration and intact consonant duration. The conclusion with respect to the discrepancy between the present study and Thorén (2004) is that the latter study included the range of durations where the listeners changed their perception of quantity category in the case of /a/ and /ʉ/, since these vowels need greater change of duration in order to be perceived as non original quantity category, and missed this range in the case of /ɛ/ and /ø/, which need less change of duration to be perceived as non-original quantity category. One can expect that the graphs from the two studies would have presented better conformity if they had both included duration values from near typical intact recording to maximal vowel and consonant duration change.

The ambiguous contribution from “correct” consonant duration to naturalness for /ʉ/, can probably be accounted for by the already damaged naturalness caused by changing of durations with intact spectral properties. In the case of /ɛ/, the listeners were probably not disturbed by “incorrect” vowel timbre, and could consequently appreciate the adjusted consonant duration more readily.

Since there is already enough evidence for the greater duration of stressed syllables in Swedish (as discussed in 1.5.2.), it can be concluded from the present study, that the duration of the post-vocalic consonant contributes to perception of quantity, and it can be assumed – in agreement with what has been reported in 1.4.3, and 1.5.2 – that it also contributes to the perception of word stress and – presumably – improved naturalness. This in turn makes it

reasonable to regard both vowel and consonant duration as important properties when learning Swedish as an L2.

### 3.3. L2-speakers' strategies to convey the Quantity contrast in Swedish – a production study

#### 3.3.1. Introduction

As discussed in 1.5.2, there are circumstances that indicate that temporal correlates to the quantity distinction are more similar over the Swedish-speaking community than spectral. Moreover, the results of the previous experiment (3.2), and the results from Traunmüller & Bigestans (1988) indicate that the postvocalic consonant duration is involved in signaling the quantity category. So far we have looked at native Swedish listeners' reaction to native Swedish speech, albeit manipulated. The present study is an attempt to create a greater variation in quantity realization than could be expected from a group of native Swedish speakers.

The variety of L2 realizations can give an indication of what phonetic/acoustic correlates function as perceptual cues to the native Swedish listener. According to Diehl & Kluender (1987: 227) "Experienced listeners make use of *all* potentially relevant cues for phonetic categories, provided these cues are detectable." Recall also the second part of H6 from Flege (1995: 239), presented in 1.2.3:

"The phonetic category established for L2 sounds by a bilingual may differ from a monolingual's if ... 2) the bilingual's representation is based on different features, or feature weights, than a monolingual's".

Examples of this phenomenon are reported by Flege (1995) and Bohn (1995). Native speakers of German, Spanish and Mandarin tried to perceive the English /i/ - /ɪ/ contrast in the word pair *beat* – *bit*, by means of spectral and/or durational cues. It turned out that native Mandarin and Spanish speakers who were inexperienced in English, tended to rely more on duration than on spectrum for the *beat* - *bit* discrimination. Native speakers of German did also rely more on duration for the same discrimination, but they were expected to do so, because German has a quantity distinction that utilizes duration to a higher degree than spectrum. Neither Spanish nor Mandarin have quantity contrasts, but seemed all the same to prefer a temporal cue for the English contrast. If native speakers of Mandarin and Spanish perceive the English contrast relying mostly on durational cues, they can also be expected to establish a representation of the contrast based on the same cues, which eventually can be expected to result in a production strategy on a durational basis (cf. Flege's H7 in 1.2.3.). The present study is interested in how

L2 productions of Swedish quantity categories will signal the intended category to native Swedish listeners, under the assumption that a substantial part of the L2 speakers will realize the quantity categories based on other phonetic cues than native Swedes. Some of the L2 speakers can be expected to have no representation of the quantity contrast, but they can still produce versions that will be perceived in a categorical way by the Swedish listeners and thus contribute to the intended variation. Although Spanish and Mandarin speakers have, as mentioned above, tended to utilize duration rather than spectrum for the quantity categorizing, other L2 speakers of Swedish may well be expected to utilize mere vowel quality for quantity categorization. The Production and perception data are thus expected to indicate whether Swedish quantity categories can be unambiguously realized using other combinations or weightings of correlates than what is common to native speakers of Swedish.

The quantity distinction in Swedish is one of the phonological contrasts that are included in the BP-approach, and consequently assumed to be of great importance for an L2 learner aiming at a listener friendly Swedish pronunciation. In the experience of the present author, there are many L2-users of Swedish who seem to have missed both the perception of the phonological contrast of quantity, and hence also fail to produce it. This often results in realizations of Swedish words, that are perceived as neither /V:C/ or /VC:/, but rather something in between. Other learners seem to have perceived the phonological contrast, but applying it only as vowel duration (e.g. Thorén 2006). Still others seem to have perceived only one or two salient spectral differences between long and short vowel allophone, e.g. the difference between [a] and [a:], and confusing the Swedish long /v:/ with an /o/, with unspecified or indistinct length category. This can be heard, and it can be seen, as some L2-users of Swedish write the Swedish long /v:/ with the å-letter, that designates Swedish /o/.

The BP-approach claims that temporal realization of quantity is a safe and reliable way to accomplish it, but what if a number of L2-learners succeed in conveying unambiguous quantity categories, applying other production strategies? The present study attempts to test whether this is possible, or if all successful L2-realizations follow the Swedish temporal pattern.

The experiments reported in 3.1 and 3.2 in the present thesis, deal with the perceptual weights of spectral and temporal correlates to the quantity distinction in Swedish. Experiment 3.2 deals specifically with the perceptual role of the postvocalic consonant. The studies are carried out by means of recordings of native Swedish speakers and systematic temporal manipulations. In order to further investigate the correlates of the quantity distinction, the present study makes an attempt to create a great variety of realizations, and this is done by letting a number of L2-speakers of Swedish do their best in pronouncing a few words that all have a counterpart consisting of the same sequence of phonetic segments, but belonging to an opposite quantity

category. The word *matta* [màt:a] ‘carpet’ is – in the light of the two previous experiments (3.1 and 3.2) – expected to be perceived as *mata* [mà:ta] ‘to feed’, if the V/C-ratio is too high, i.e. if the relative vowel duration is too long and/or the relative consonant duration is too short. Consequently, the word *vila* [vì:la] ‘to rest’ can be perceived as *villa* [vìl:a] ‘villa’, if the /i/ has too short a duration and the /l/ has too long a duration.

As is mentioned before, and is discussed in 1.4.2, the Swedish quantity distinction is traditionally described as “long or short vowel” but it has been understood for a long time, that there is a complementary relationship between the vowel and the postvocalic consonant, thoroughly described in 1.4.3 and 1.5.1. This complementary relationship exists to different degrees in different varieties of Swedish (Reuter 1971; Gårding 1974; Strangert & Wretling 2003; Schaeffler 2005), but does not seem to be completely absent in any variety.

The experiments 3.1 and 3.2, as well as Jonasson & McAllister (1972) show that the V/C ratio is relevant to the listener’s categorization in /V:C/ and /VC:/ respectively. This measure also accounts for variation in speaking rate. From Elert’s (1964) production data, it is clear that the V/C-ratio is higher for a /V:C/-word than for a /VC:/-word, provided that the sequence of vowel and consonant phonemes is the same, and that the VC-sequence of the words have the same distance to word- phrase- or utterance final position. With different vowel and consonant phonemes come different inherent durations, that will affect the V/C-ratio. An open vowel has a greater inherent duration than a closed vowel, and a voiceless consonant usually has a greater inherent duration than its homorganic voiced counterpart, and in addition, a voiceless consonant tend to shorten the preceding vowel, while a voiced consonant tends to lengthen the preceding vowel. This is clear from the results of Elert (1964). Recall also the discussion in 1.6, and the well known condition that quantity is realized only in stressed syllables, and as mentioned in 1.4.3, the extra duration that follows from focus accent, falls mainly upon the phonologically long segment, thus increasing the difference in V/C-ratio between /V:C/ and /VC:/ words.

Traunmüller & Bigestans (1988) found that the vowel duration at the boundary between /V:C/ and /VC:/ words, is well correlated with the duration of the whole utterance, when manipulated words were presented in a carrier phrase at varying speaking rates. Their results showed corresponding but somewhat lower correlation between consonant duration and utterance duration, and the latter correlation was also depending on which consonant phoneme was tested. Their experiment included one vowel phoneme /ε/, and three consonant phonemes; /n/, /s/ and /t/. Their findings are in agreement with the claims of Diehl & Kluender (1987:227), cited in 1.6. The relative measures mentioned here and above are mutually dependent, since increase in the duration of a vowel will have an effect on the duration of the word and of the utterance of which it is a part. Increased vowel duration will conse-

quently result in a higher V/C ratio, a higher vowel over word ratio and a higher vowel over utterance ratio, and simultaneously contribute to a lower consonant over word ratio and consonant over utterance ratio. Increased duration of a postvocalic consonant will affect relative measures of vowel duration in a corresponding way. So just because the postvocalic consonant is not mentioned in the measure that best correlates to native listeners' quantity categorization, it may still contribute to the listener's categorization into /V:C/ or /VC:/.

As is evident from 1.5.2 and experiment 3.1, the quantity contrast has – in addition to temporal correlates – also spectral correlates that are substantial in some vowel phonemes and negligible in some, see 3.1.4. It is also one of the central assumptions of BP (cf. 1.5.2.), that temporal correlates of the quantity distinction are generally more valuable to the L2-learner of Swedish, than spectral.

As can be concluded from Hadding-Koch & Abramson (1964), Behne et al. (1997) and experiment 3.1, vowel duration and spectrum have different weights or degrees of importance, as perceptual cues to the quantity contrast, for different vowel phonemes. The results also indicate that the relative duration of the vowel is the main acoustic correlate to the Swedish quantity contrast, for all vowel phonemes except /ʉ/, although spectral properties assist as perceptual cues for /a/ (Behne et al. 1997). Recall the rank order in 3.1.4, /ɛ/ /ø/ >> /o/ >> /i/ /a/ >> /ʉ/ that was concluded from the opening study. Behne et al. (1997) and Johansson (1981) show that quantity categorization for /a/ is based on a combination of temporal and spectral cues.

As is concluded from the previous section 3.2.4, the relative duration of the postvocalic consonant also seems to assist the listener in perceiving quantity category in cases where the vowel has durations in between typical values for either /V:C/ or /VC:/.

A number of L2-speakers with different L1's and different levels of proficiency in Swedish, are expected to produce a great variety of temporal and spectral realizations of Swedish words that are parts of minimal pairs with respect to quantity, e.g. *vila* [vì:la] 'to rest' – *villa* [vìl:a] 'villa'. This variety could hopefully provide a good share of unambiguous versions, presenting different utilization of temporal and spectral correlates.

The present study attempts to find out whether a number of L2-speakers of Swedish, with expected great variation in quantity realization, can produce unambiguous versions with other temporal structure than a group of native Swedish speakers. The main purpose is to study the quantity contrast correlates in those speakers who are successful in conveying a quantity category, and to compare the use of correlates to some native Swedish speakers. The L2 realizations of Swedish quantity category will be evaluated by native Swedish listeners, with respect to quantity category and ambiguity. Special attention will be paid to how the duration of the postvocalic consonant will be utilized in unambiguous versions.

McAllister et al. (2002) studied L2-users' realization of the Swedish quantity contrast, and formulated a "Feature Hypothesis", that predicts that a phonological contrast in a L2 is more difficult to acquire if it builds on a phonetic property not utilized in the L1. The phonetic property in the case of quantity is assumed to be duration. Their investigation shows that persons with Estonian as L1 manage better in both perception and production of the Swedish quantity contrast, than do native speakers of American English, who in turn manage better than native speakers of South American Spanish. This is in good agreement with how the respective L1's utilize duration: Spanish has no phonological quantity, English has a contrast between tense and lax vowel in stressed syllables, that also entails durational differences, and Estonian has separate vowel- and consonant quantity, with three distinctive degrees of length.

Questions in the present study

- Will unambiguous L2-productions utilize vowel duration and consonant duration in similar ways as native Swedish ones? Will there be unambiguous L2-realizations falling outside the temporal patterns of Swedish L1-realizations, utilizing spectrum to a higher degree? And correspondingly; will there be unsuccessful L2-realizations falling inside the temporal range of Swedish L1-realizations?
- Will the native Swedish listeners, in their categorization of L2-realizations in the present study, relate segment durations to the durations of whole utterances, thus agreeing with the results of Traunmüller & Bigestans (1988), and the claims of Diehl & Kluender (1987)?
- Will native Swedish productions and successful L2-productions give support for the hypothesis of pedagogical elasticity, mentioned in 1.5.3.2?
- Will the L2-speakers have a better command of the Swedish quantity contrast if their L1 utilizes phonological quantity?
- Will the speakers' background – level of education, time of residence in L2 environment, or age at arrival in Sweden – have any impact on how well they succeed to convey the Swedish quantity contrast?

Since native Swedish speakers seem to use duration as the dominant correlate to the phonological quantity for at least seven out of nine vowel phonemes, and make significant temporal changes between long and short allophone also for /ʉ/ (cf. table 1.4) that is considered the most spectrum-dependent phoneme with respect to long-short difference, the L2-speakers are expected to utilize duration as the main strategy to convey the phonological quantity. Moreover, the L2-speakers whose L1 utilizes phonological quantity are expected to be more successful in conveying the Swedish quantity contrast, than those L2-speakers whose L1's do not.



### 3.3.2 Method

#### 3.3.2.1. Material

33 speakers of Swedish as L2 were recorded when reading 6 Swedish words, *gratis* [gró:tis] ‘free of charge’, *matta* [mà:t:a] ‘carpet’, *vila* [vì:la] ‘to rest’, *sillen* [sí:l:ən] ‘the herring’, *mäta* [mè:ta] ‘to measure’ and *etta* [èt:a] ‘the figure 1’. The words were read in the carrier phrase *Det var ..... jag menade* ‘it was ..... that I meant’. The recording of 10 native Swedish speakers reading the same words was done for reference. Each test word has a counterpart with the same sequence of phonemes but with opposite length relations in the VC-sequence. If e.g. *gratis* [gró:tis] ‘free of charge’ is pronounced with short /a/ and long /t/, it will be perceived as *grattis* [grát:is] ‘congratulations’.

In addition to written text, pictures and explanations were presented to the L2-speakers, to assure that they – since they represented many different levels of command of Swedish – understood the meaning of the word they were producing. They were, however, not given any hints concerning the pronunciation of the words. The L2-speakers were instructed to read the sentence in a continuous way. This succeeded mostly, but occasional recordings contain pauses.

The words were recorded in a sound treated room in the present author's home and in a studio at the University of Stockholm, using a Røde NT3 condenser microphone and a Sony MZ-N710 mini-disc player. Out of the totally 198 recorded words, 19 were omitted due to insufficient quality. 179 words in carrier sentences were measured and presented to native Swedish listeners.

The choice of /a/, /ɛ/ and /i/ is the result of the intention to include vowel phonemes with different degree of spectral difference between long and short allophone. The spectral difference between short /ɛ/ and short /e/ is neutralized in most central Swedish varieties, explaining why *mäta* and *etta* can be used as opposite quantity categories of the VC-sequence. /a/ was chosen due to its relatively large spectral difference between long and short allophones, and has the advantage of being part of the vowel inventory of practically every language in the world. The Swedish /ʉ/ has an even greater difference in this respect, but experience tells us that many L2-users have great difficulties producing an acceptable /ʉ/-segment at all, and/or distinguishing it from /u/, and sometimes from /y/.

/i/ was chosen because of its intermediate position in the dependence of duration-spectrum to signal long/short allophones, and also because of a typical (in central Swedish) palatal phase at the end of the long allophone [ij].

/ɛ/ is the vowel phoneme with the least spectral difference between long and short allophone, and is consequently most dependent on duration for signaling of long/short allophones.

Genuine minimal word pairs were avoided in order to conceal a phenomenon that represents a great difficulty to many L2-users, and thus prevent them from feeling tense during the recording. Some of the speakers probably realized the purpose of the test words, but nevertheless all speakers but one seemed to pronounce the test phrases without nervousness.

### **3.3.2.2. Speakers**

The speakers were between 17 and 61 years of age, and had been living in Sweden from 9 months to 24 years. They had 18 different L1's, 6 of which utilized some kind of phonological quantity. All but 1 speaker – who came to Sweden at the age of 7 – had all started their learning of Swedish as adults or nearly adults.

### **3.3.2.3. Segmentation and measures**

The words were segmented and measured in Praat (Boersma & Weenink 2001). Those cases of segmentation where words ended in /a/, having no clear boundary to the following approximant, are accounted for in 2.1. Utterance duration, word duration, first vowel duration and duration of the post-vocalic consonant were measured, and the following ratios were calculated: V/C ratio, vowel duration over word duration (V/W), vowel duration over utterance duration (V/U), consonant duration over word duration (C/W) and consonant duration over utterance duration (C/U). Correlations were calculated for these measures and the number of responses as “intended quantity category”. For unambiguous L2-realizations that fell outside the temporal patterns of the native Swedish realizations, two spectral measures were made for /a/ and /i/ respectively. In words containing /a/ as their first vowel, F1 and F2 frequencies (Bark) were measured and compared to the productions of the native Swedish speakers. For /i:/, the occurrences of a palatal fricative phase was examined. The typical spectral correlates of the /ij/-diphthong is – in addition to audible impression – a steep decrease in the intensity curve after roughly 25% of the vowel duration, combined with a distinct rise in F3, which falls again after some 50% of the vowel duration, the same point in time where intensity begins to increase after the fall. This, in combination with a weakening of F2 in the mid part of the vowel, serves as criterion for a diphthongized [ij]. Correlations were also calculated between success in conveying quantity categories and background variables.

### **3.3.2.4. Categorizing task**

The words in carrier phrases were presented in random order for categorization to 10 native Swedish listeners, who marked the words as ‘/V:C/- word’, as ‘/VC:/-word’, or ‘undeterminable’. The listeners came from different regions in Sweden, and none of them had any hearing deficiencies that affected their perception of normal speech.

### 3.3.3. Results

#### 3.3.3.1. Temporal measures

The temporal realizations are shown in figure 3.3.1 by means of the V/C-ratio, and in figure 3.3.2 by means of vowel duration over utterance duration (V/U). By "successful" or "good" productions is meant that all 10 native Swedish listeners perceived the word as "intended quantity category". Productions that received 0 responses as "intended quantity category" are counted as "unsuccessful" productions. One important finding with respect to the native Swedish listeners, is that they do not perceive the quantity categories uniformly. There are several cases where some of the native Swedish listeners have responded "intended category", some have responded "undeterminable" and some "opposite category", e.g. *mata* [mɑ:ta] 'to feed' in stead of *matta* [mɑ:t:a] 'carpet'. Criteria for categorizing into "successful" and "unsuccessful" can be discussed, but adopting the present, very strong criteria, will safeguard that the categories will contain only unambiguous realizations, although a number of "acceptable" realizations will be omitted.

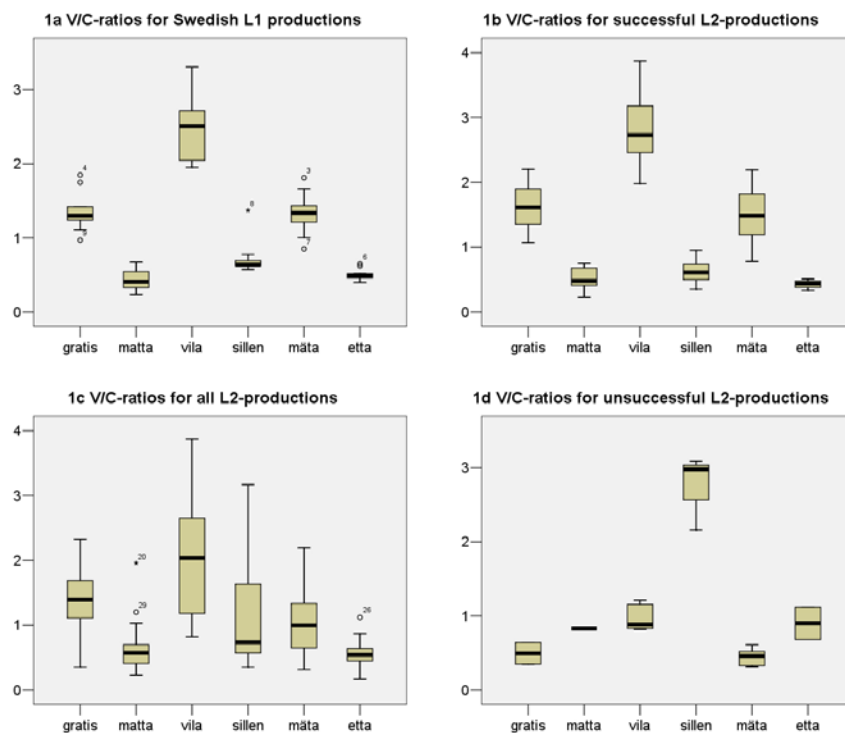


Figure 3.3.1 a-d. Distribution of V/C-ratios for the 6 test words. a) native Swedish speakers, b) successful L2-productions, c) all L2-productions, and d) unsuccessful L2-productions.

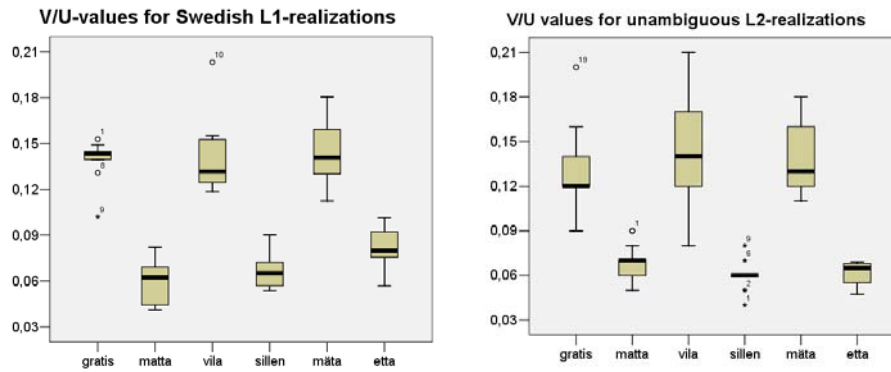


Figure 3.3.2. Relative vowel durations (vowel over utterance duration) for all test words. Left panel: Native Swedish speakers, right panel: unambiguous L2-versions.

The L2-speakers as a group showed great dispersion and overlapping, as can be seen in figure 3.3.1c, while the successful L2-productions and the productions of the native Swedes show an almost identical pattern i.e. higher V/C-ratios for /V:C/-words than for /VC:/-words, and hardly any overlapping between word pairs with the same VC-sequence. This is in agreement with the results of Elert (1964). The unsuccessful productions in 3.3.1d show an almost opposite pattern, compared to the successful productions, i.e. lower V/C-ratios for intended /V:C/-words and higher for intended /VC:/-words. V/C-ratios do not tell us whether the VC-sequences follow a complementary pattern or not, since variations in vowel duration or consonant duration separately may influence the ratio. Figure 3.3.2 shows relative durations of short and long vowels in all test words, by means of the Vowel over utterance duration ratio. The figure shows that native Swedish speakers separate all /V:C/-words from /VC:/-words with this measure, with a small “safety-zone” between the categories, whereas the L2-versions show a large zone between long and short /ε/, and small overlapping in the case of /a/ and /i/.

The box plots in figure 3.3.3 indicate that successful L2-realizations follow the same patterns as native Swedish speakers with respect to both vowel and consonant durations, although L2 speakers produce somewhat shorter relative durations in the case of phonologically long segments, and correspondingly shorter relative durations in the case of phonologically short segments, thus maintaining V/C ratios that are on a par with the Swedish L1 realizations, as shown in figure 3.3.1. The difference between L1 and L2 speakers with respect to relative duration of all phonological long segments is only just significant ( $p = 0.046$ ), and the corresponding difference with respect to short segments is only just below significance ( $p = 0.056$ ).

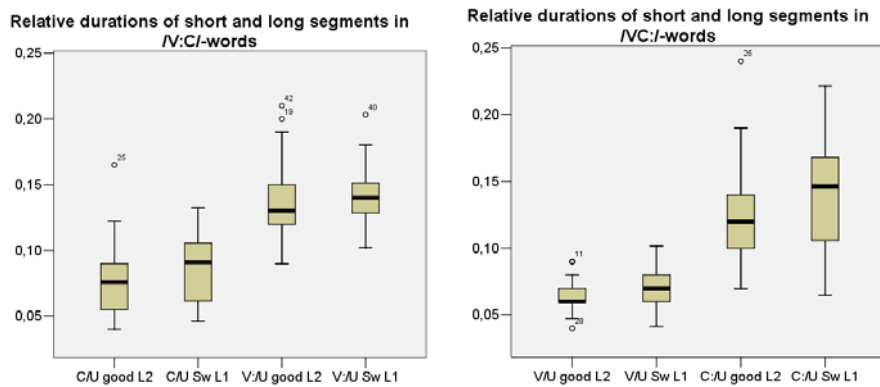


Figure 3.3.3. Left panel shows relative durations (segment/utterance) for phonologically short and long segments in /V:C/-words. “Good L2” means that all 10 Swedish listeners perceived the word as intended quantity category. Right panel shows corresponding values for /VC:/-words.

The degree of complementary duration can also be illustrated by the V:/V and C:/C-ratios, and those are given in table 3.3.1 below. It is clear that the successful L2-productions apply both vowel and consonant durations according to a complementary pattern. There were not enough speakers with zero responses for intended quantity category, to make the corresponding calculations for “unsuccessful” productions.

Table 3.3.1. V:/V and C:/C-ratios (m) for native Swedish speakers and L2-speakers respectively. n = number of productions, m = mean value, and sd = standard deviation. Mean values are derived from within-speaker ratios.

	<b>Native Swedes</b>	<b>Successful L2-productions</b>	<b>Native Swedes</b>	<b>Successful L2-productions</b>
	<b>V:/V</b>	<b>V:/V</b>	<b>C:/C</b>	<b>C:/C</b>
<b>n</b>	<b>30</b>	<b>23</b>	<b>30</b>	<b>23</b>
<b>max</b>	<b>3.08</b>	<b>4.13</b>	<b>2.64</b>	<b>2.91</b>
<b>min</b>	<b>1.38</b>	<b>1.37</b>	<b>1.18</b>	<b>0.81</b>
<b>m</b>	<b>2.01</b>	<b>2.38</b>	<b>1.61</b>	<b>1.6</b>
<b>sd</b>	<b>0.39</b>	<b>0.77</b>	<b>0.36</b>	<b>0.51</b>

Table 3.3.2 below shows the correlations between number of responses for “intended quantity category” and the different suggested measures. The vowel-based correlations for /VC:/-words are negative, since lower ratios yield more /VC:/-responses. Vowel duration divided by the duration for the whole utterance, is best correlated to number of responses for intended quantity category. Correlations between relative consonant duration and number

of responses for intended quantity category, is lower than for vowel durations, but still significant.

Table 3.3.2. Correlations between 4 different temporal measures and number of responses for "intended quantity category" for each utterance.

/V:C/-words			/VC:/-words		
V/C	r = 0.61	p < 0.01	V/C	r = - 0.60	p < 0.01
V/W	r = 0.62	p < 0.01	V/W	r = - 0.37	p < 0.01
V/U	r = 0.74	p < 0.01	V/U	r = - 0.67	p < 0.01
C/W	r = -0.62	p < 0.01	C/W	r = 0.31	p < 0.01
C/U	r = -0.46	p < 0.01	C/U	r = 0.4	p < 0.01

Another way of testing the suggested ratios with respect to how well they agree with native listeners' categorization into /V:C/ and /VC:/-words, is to find out how they temporally separate unambiguous realizations from unsuccessful realizations. The best vowel-based measure should have no or minimal overlapping between the lowest value for unambiguous /V:C/-words (including L1-versions) and the highest value for unsuccessful /V:C/-words. The best ratio should, of course, also separate unambiguous /V:C/-words (including L1-versions) from unambiguous /VC:/-words.

Within each word pair with the same VC-sequence, the quantity categories were differentiated by all measures (see table 3.3.2.) except vowel duration divided by word duration (V/W). This is true for the native Swedish speakers, and with isolated cases of overlapping in the successful L2-productions. Vowel duration divided by utterance duration (V/U) separated the quantity categories for both native Swedes and successful L2-productions. None of the ratios/measures could separate the quantity categories for the whole material, but the V/U-ratio showed the least degree of overlapping.

This is in agreement with the fact that the V/U-ratio was best correlated to the number of responses for intended quantity category.

### 3.3.3.2 Realizations that do not agree with native Swedish boundaries

In agreement with correlations in the present study, and with the findings of Traunmüller & Bigestans (1988), the V/U-ratio is regarded as a main measure. Unambiguous realizations that are not within the range of native Swedish V/U values, as well as unsuccessful realizations that fall within the V/U range of native Swedish speakers are reported here. This means below a certain V/U value for /V:C/-words, and above a certain value for /VC:/-words.

For the tests word *gratis* [gró:tis], we can observe an unambiguous L2-realization that is slightly under the V/U value, which is minimum for L1-

speakers. At the same time the V/C and V/W values are well above the minimum among L1-speakers. A closer look at the utterance shows a hesitation pause of 300 ms before, and a 140 ms pause inserted after the test word, which lengthened the utterance considerably. The word final and prepausal /s/ of *gratis* can also be expected to be subject to final lengthening. This results in lower values of durations related to the duration of the whole utterance.

In the case of the test word *mäta* [mè:ta], there is one unambiguous L2-realization that lies under the minimum V/U-value for Swedish L1-realizations, but again, the speaker makes a rather long pause (430 ms) after the test word. V/C and V/W for the mentioned version are well above L1-minimum. One speaker is well above the V/U and V/W minimum, but below minimum for /V:C/ with respect to V/C and above maximum for C/U, but this realization still has 10 responses for intended quantity category. This case of lengthening of a phonologically short segment is seen as an outlier in figure 3.3.5 below.

The same happens in *vila* as in *mäta*. One unambiguous L2-realization shows lower V/U-values than minimum for L1-productions. Also this time, there is a hesitation pause before the test word. The V/W-value is only just below the corresponding L1 minimum, which indicates that the pause does not account for all of the relative shortness in the vowel. The V/C value however, is substantially above the L1-minimum for this test word, and the speaker fulfills at least two criteria for the palatal diphthongization of the long /i/ (see 5.2.3).

As for *matta*, there are three unambiguous L2-realizations which have V/U values higher than the maximum for L1-speakers; V/U 0.09 whereas no L1 realization exceeds 0.08. These three realizations are also higher on V/C, and V/W ratios than maximum for Swedish L1-realizations. They do, however show C/U-ratios above or same as minimum L1-realizations. It should also be noted that they have V/U values lower than minimum L1-values for *gratis*. The mentioned realizations are produced by three male L2 speakers, and the spectral properties of the L2-realizations are compared to the corresponding properties of five male L1-speakers, producing *matta*, and *gratis*, the latter of which is the word of opposite quantity category. Figure 3.3.4 shows this comparison in a F1-F2 vowel space, and it seems clear that the three L2 productions in question, have spectral qualities that agree well with L1-values for short /a/-allophone. All successful L2-versions of *matta* have C/U-ratios above or same as minimum among L1-speakers. For *etta*, all successful L2-versions had V/U-values below or same as maximum for L1-speakers, and C/U-values above minimum for L1-speakers. The same is true for *sillen*. This shows that all successful L2-versions had relative consonant durations within the range of L1-speakers.

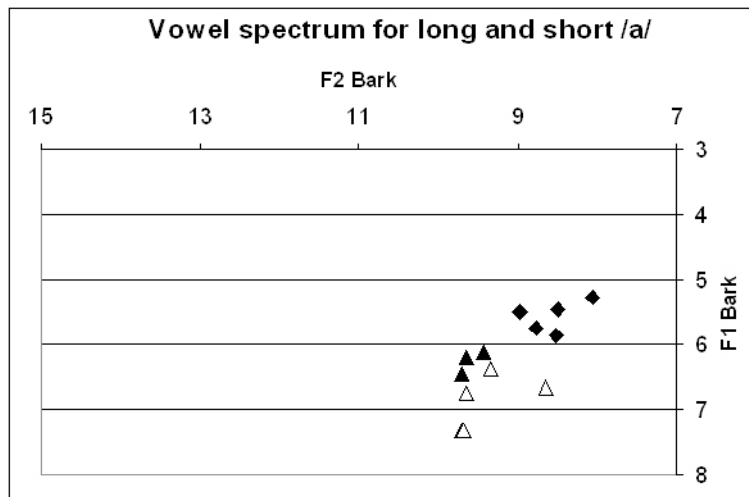


Figure 3.3.4 Vowel spectrum for long and short /a/ produced by male speakers. Black diamonds are L1-realizations of *gratis*, open triangles are L1-realizations of *matta*, and filled triangles are L2-realizations of *matta*, with high V/U-ratio.

### 3.3.3.3. The effect of speaker background

Of the background variables, two seemed to have significant influence on how well the L2 speakers managed to convey the Swedish quantity categories, namely whether the speakers' L1 used phonological quantity or not, and length of education (in years). A one way analysis of variance (ANOVA) yielded  $p < 0.05$  when comparing the "mean successful production scores" for the group of L2-speakers whose L1 used phonological quantity, with the group of L2-speakers whose L1 did not. When tentatively adding the productions of one Estonian speaker – whose recording was technically insufficient for duration measures – to the comparison, the p-value went down to just above 0.01. This version was not presented to the judge of 10 native Swedish listeners, but was judged as totally unambiguous by the present author. Correlation between "mean successful production scores" and number of years in education yielded  $r = 0.31$  which in turn yields  $p < 0.05$  in a one-tailed significance test.

The remaining 2 background variables; age of arrival in Sweden and length of residence in Sweden, showed correlation coefficients 0.03 and 0.07 respectively.

### 3.3.4. Conclusion and discussion

The results of the present study show that successful L2-productions present a very similar temporal pattern compared to the ones made by the native Swedish speakers, including complementary consonant duration. The results



also show that the unambiguous realizations of the quantity categories /V:C/ and /VC:/, can not be separated in the whole material by the temporal measures used in this study, but only for each word pair, containing the same VC-sequence. This is not surprising considering the different inherent durations to be found in different vowel and consonant phonemes, and the impact the type of postvocalic consonant may exert on the preceding vowel. To capture the perceptual quantity categories instrumentally, we would need some kind of index value for the respective inherent durations of different vowel phonemes, and probably also for combinations of vowel phonemes plus following consonant phoneme, since there is also an interaction between vowel type and consonant type with respect to durations in these both segments, as mentioned in 3.3.1.

The fact that the vowel/utterance (V/U) duration ratio was the best measure to distinguish the quantity categories – with respect to both perception and production – agrees with the results of Traunmüller & Bigestans (1988), which was mentioned in the introduction. It is also in agreement with Diehl & Kluender (1987), which was mentioned in 1.6, i.e. the listener makes use of all possible cues for the categorization, and this includes relating e.g. vowel duration to entities above segment and syllable level. The V/U measure has now been tested not only as boundary between /V:C/ and /VC:/ words, but also as a parameter correlated to the likelihood of words being categorized as either /V:C/ or /VC:/. One reason for the somewhat lower correlation between vowel over word ratio could be the difficulty described in 2.1, to securely identify the boundary between the final vowel of the test words *etta*, *matta*, *mäta*, *vila*, and the following palatal approximant. The same uncertainty zone, e.g. 10 milliseconds, makes up a greater share of the word duration than of the duration of the whole utterance. Furthermore, the problem with the mentioned transition is not present in the case of vowel over utterance ratio. A carrier phrase that had e.g. a voiceless obstruent following the test word might have yielded other correlations. On the other hand, some cases of pauses inserted before or after the test word, as described in 5.3.2, have occasionally resulted in vowel duration over word duration (V/W) being a better measure.

Did the present results give any support for the hypothesis in 1.5.3.2, that phonologically long segments would have more generous durational limits than phonologically short segments, once above a critical value for quantity category? The phonologically short segments do of course have natural limits – a minimal duration to be perceived at all, and an upper limit to maintain the quantity category. As mentioned in 1.4.3, phonologically long segments receive extra duration in focal position, while other segments in the focused words are not affected by the focal position. Hypothetically the phonologically long segments could have unlimited duration, since the extra length could not harm the signaling of quantity category, but it could, of course be expected to harm the naturalness. Still variation in speaker style and com-

mand of Swedish pronunciation could make it likely that phonologically long segments showed a greater dispersion in duration than phonologically short sounds. The boxplots in figure 3.3.3 gives an implication that this may be the case. Figure 3.3.5 shows all phonologically long segments from L1-realizations and successful L2-realizations respectively and the corresponding groups of phonologically short segments. The phonologically long seg-

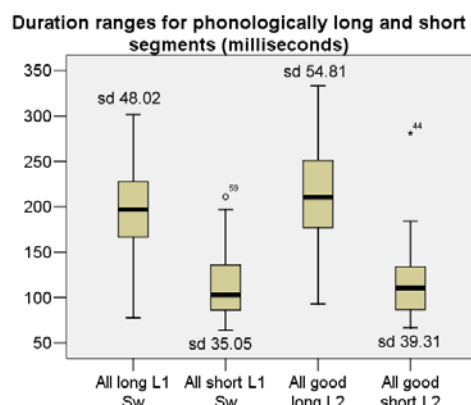


Figure 3.3.5. Absolute durations in all phonologically long segments and all short segments, in Swedish L1 productions and unambiguous ('good') L2 productions. Standard deviations above or under each box plot.

ments show higher values and a greater dispersion, and consequently higher standard deviations, a result that gives some support for the hypothesis of phonetic elasticity. The outlier (between 250 and 300 ms) among "good short L2" in the rightmost plot in the figure is commented on in 3.3.3.2.

Most of the L2-speakers produced a spectral difference between long [v:] and short [a], and all unambiguous realizations of *gratis* were made with back [v:]. This can be natural provided that the differences could be audible to many L2-users, but it can also be possible that the Swedish [v:] can be perceived as /o/, rather than an /a/ for a person with an L1 containing 5 vowel phonemes, which is the most frequent vowel inventory (Maddieson 1981). We hypothesize that this person is not used to distinguishing more than three degrees of openness in the back vowel space, and the difference between long [v:] and short [a] can thus be expected to be perceived as phonemic by some of the L2-users. In the present author's experience, there are quite many L2-learners of Swedish, who write the "dark" long [v:], with the letter 'å' that is intended only for /o/, as mentioned in 3.3.1. Three cases of high V/U values for the word *matta* may have been saved by the spectral properties of the vowel. One could otherwise expect that some of the Swedish listeners would have responded "undeterminable" as a result of too long relative vowel duration. The three mentioned cases cannot be accounted for by utterance lengthening pauses, since pause-induced increase in utterance duration would have caused a decrease in V/U value. It is also possible that

their C/U ratios contributed to their unambiguity, since they had C/U values same as or above native Swedish versions. C/U is significantly correlated to number of both /V:C/ (negatively) and /VC:/ (positively) responses (cf. table 3.3.2).

The majority of the successful L2-productions did not apply the typical fricative palatalization of long /i/, except one version of *vila* that had a V/U value lower than minimum for L1-speakers. This L2-version did however contain rather long hesitation pauses, which is likely to distort the V/U-ratio. She had V/W-ratio nearly within the L1-range, and V/C-ratio well above the L1-minimum. The fact that this spectral signal was so sparsely utilized, could possibly be an example of findings reported by Cebrian (2007) and Bohn (1995), which indicate that durational differences are more easily adopted by L2-learners, than spectral differences. The fact that some Swedish dialects have minimal or no spectral difference between long and short vowel allophone, even for /a/ (Finland-Swedish, Reuter 1971, and Gotland-dialect, Eriksson 2004) and /u/ (Finland-Swedish and South-Swedish area between Helsingborg and Lund, which was discussed/mentioned in 1.5.2.) can also indicate that duration is the correlate to quantity, that is common to all varieties of Swedish.

Of the background variables mentioned in 3.3.1 and 3.3.3, level of education and the L1-use of phonological quantity were those that yielded a significant correlation and a significant difference between speaker groups respectively. Length of residence in the L2-community did not show significant correlation with “mean successful production scores”. This is consistent with the results of McAllister (2001), where it was shown that native English learners of Swedish did significantly better when trying to acquire the Swedish quantity contrast, than did native Spanish learners. The difference in “length of residence” within or across groups did not correlate with “success in acquiring the Swedish quantity distinction. These results, including the present, supports the “feature hypothesis” (McAllister et al. 2002) mentioned in 3.3.1 above. One could possibly expect that the great differences in “length of residence” in the present study (3 months – 24 years) would have shown a significant influence, but they did not. In the discussion of McAllister (2001), it is pointed out that certain studies have come to the similar conclusions (e.g. Moyer 1999), but also that other studies (e.g. Flege 1988, Purcell & Suter 1980) have concluded that length of residence in the L2 environment does have a positive effect on the L2 pronunciation result. There are no data available in the present study concerning quality and amount of L2 input for the speakers of Swedish as L2. The two latter variables were concluded as significant for phonetic development in adults, by Flege & Liu (2001).

There are difficulties drawing clear boundaries between languages that utilize phonological quantity, and languages that do not. English and Dutch are listed in Garlén (1988) as not having phonological quantity, whereas

McAllister et al. (1999) and Ladd et al. (1999) respectively, describe them as having phonological quantity. One could also expect that speakers of languages that have obvious durational differences, although in complementary distribution, such as French and Persian, would have an advantage learning the Swedish quantity distinction, compared to other languages, and this proved correct, as adding speakers of French and Persian to the group of “quantity languages”, the significance with respect to successful productions, improved from  $p < 0.05$  to  $p < 0.01$ .

A cautious conclusion with respect to pedagogical consequences of the present result, is that duration is the most safe way of realizing the Swedish quantity distinction, since all unambiguous productions were “temporally correct” in some respect i.e. had relative durations similar to those of the native Swedish speakers. It is also of interest that not only V/C-ratios and V/U-ratios were native-like in the successful L2-realizations, but also that these applied complementary consonant length. Both V:/V and C:/C ratios for successful L2-realizations were, as shown in table 3.3.1, quite near those of native Swedish speakers. This should by no means defend a teaching strategy where spectral differences are totally neglected, but it has significance with respect to what is regarded as necessary and optional respectively, and in what cases great efforts are made when the learner has difficulties.

One source of uncertainty in this study could be that other native Swedish listeners did not judge the productions made by the native Swedish speakers. It was taken for granted that these native speakers would realize the quantity categories in an unambiguous way. This assumption can be somewhat premature, since the Swedish listeners showed great variation in their judgement of the L2-productions. This in turn, indicates that the boundary between the quantity categories, based on mere acoustic cues, may not be uniform for native Swedish speakers, and it can well be assumed that the semantic context plays an important role when the native language user perceives quantity category.

### 3.4. L2-speakers’ strategies to convey the word stress contrast in Swedish – a production study

#### 3.4.1. Introduction

The basic arguments for the interest in acoustic correlates of stress are presented in 1.5.2. The present study can be seen as a test concerning the accuracy of the BP claim, that stress in Swedish must be realized by means of increased syllable duration, an increase assumed to be perceived as extra length and prominence in the pedagogic setting. This extra duration is also

said to be associated to either vowel or consonant of the stressed syllable, which in turn is assumed to contribute to maintaining the correct quantity category. Will this prove valid when native Swedes try to categorize a great variety of L2-realizations with respect to word stress, or will other strategies result in unambiguous versions?

Just like in study 3.3 (previous section), much can be learnt about a phonological contrast by creating a large variation in its realization. One way of doing this is letting L2-speakers apply the contrast of interest. This study is concerned with Swedish word stress, and more specifically with its temporal acoustic correlates. Stress is a highly abstract concept, but is always associated with perceptual prominence. In Swedish, stress is always accompanied by greater syllable duration (Fant & Kruckenberg 1991b, Strangert 1985), Eriksson (1991), and the extra syllable duration can be seen as one side of the realization of quantity, i.e. a stressed syllable must be either of the /V:C/ or the /VC:/-category. In higher degrees of sentence-related prominence, i.e. accentuated and focal position, the extra duration is also accompanied by a pitch movement. Higher prominence degrees like focus, or sentence accent, which are mostly described in terms of the tonal gesture (c.f. table 1.3.), also result in additional syllable duration (Bannert 1979a, Heldner & Strangert 2001).

The present study has, as the previous one, bearing on Flege's (1995) Speech Learning Model H6 and H7, in that an L2 contrast or category can be perceived by the learner, but be based on other correlates or other weights of correlates than what is the case for native speakers of the target language (cf. the discussion in 1.2.3 and 3.3.1). Although Fant & Kruckenberg (1994) conclude that duration is the most robust correlate to stress in Swedish, stress in Swedish has at least three phonetic correlates, depending on phonological prominence on sentence level:

- Temporal – Stressed syllables have longer duration than unstressed syllables. One function of this extra duration is required to signal quantity category.
- Tone – sentence related prominence levels “accentuated” and “focused” (Bruce 1977), are assumed to entail a tonal fall, early for accent 1 and later for accent 2, and in focal position there is a tonal rise in addition to the word accent fall. The rise occurs in the stressed syllable in accent 1 words, and in the post tonic syllable in accent 2 words. It was found, e.g. by Engstrand (1995), that the word accent fall in accent 1 words does not occur as regularly as the fall in accent 2 words, meaning that only accent 2 words have a characteristic tonal pattern, while accent 1 words receive their tonal patterns from the sentence intonation.
- Spectral – in addition to the fact that all 9 Swedish vowel phonemes can contrast only in stressed syllables, the realization of vowel phonemes are generally closer to a “schwa” [ə] in unstressed syllables, and consequently

occupying the peripheral parts of the vowel space in stressed syllables. Engstrand (1988:1863) showed that “...stressed vowels (/i a u/) displayed narrower oral tract constrictions than unstressed vowels...” They are tenser, to apply a concept often used to describe the English closest equivalent to the Swedish quantity contrast.

There may be some 20 or 30 minimal pairs in Swedish word stress, and some of them can be regarded as semi-minimal, since the trochaic member is often an accent 2 word, a category that is not possible for a iambic word, see e.g. Bruce (1977, p18). Thus it can be argued that the word pair is distinguished by more than one distinctive feature, e.g. *bànan* [bò:nən] ‘the course’ – *banán* [benó:n] ‘banana’. It is however not primary for the risk of confusion, that word stress in Swedish is considered important in an L2-learning perspective, but because of general problems with intelligibility and recognition of words, when the temporal organization is distorted. Recall the findings of Bannert (1986) and Field (2005), reported in 1.3.2.

In order to test whether listeners utilize perceptual cues outside the word, a carrier sentence will be used in this study, just as in the previous study, although one could expect that the word stress is mainly a contrast on word level, i.e. listeners will search for cues of relative prominence of the syllables within the word, by comparing the prominence of the word-internal syllables.

In order to test whether syllable duration is a robust correlate to perceived stress also among a variety of L2-realizations, there must be a definition of syllable boundaries, that is both perceptually relevant and at the same time robust in terms of salient measurement points. This study follows the suggestion of Fant & Kruckenberg (1989), that regards the VC-sequence as the syllable “thus avoiding to split a consonant into two parts and securing the integrity of /V:C/ and /VC:/ units which we know to be comparable in duration and to be the main carriers of stress induced duration” (Fant & Kruckenberg 1989:11). It can be argued that a phonologically long consonant, e.g. [l:] in *kallas* [kàl:əs] ‘is called’, should be split into its components – according to the idea of gemination cf. 1.5.3 – when doing a correct syllabification, but this raises problems as to where along the consonant duration, this boundary should be defined. If the idea of a long postvocalic consonant contributing to the syllable length is to be defended, then a substantial part of the postvocalic consonant has to belong to the first syllable. To define the VC-sequence as the temporal unit seems reasonable in this study, although it violates the classical way of dividing words into syllables. In the present study *ja'pan* ‘Japanese’ and *'filen* ‘the file’ are parsed into *jap.an* and *fil.en* instead of the classic *ja.pan* and *fi.len*. The word initial consonant is not included in the syllable measure. Advantages with the VC-sequence as syllable duration gauge:

- It yields robust measurement points coinciding with segment boundaries
- It divides all of the words in the present study into two VC-sequences that can be compared in duration, provided that we exclude the word-initial consonant.
- It allows a phonologically long consonant to belong entirely to the syllable it is assumed to lengthen.

In order to test the role of the postvocalic consonant as carrier of the extra duration induced by stress, there are two words '*kallas* [kàl:əs] 'is called', and '*racket* [rák:ət] 'racket', in which the speakers, both L1 and L2, have the possibility to lengthen a postvocalic consonant by means of the occlusion for /k/ in *racket*, and lengthening the entire /l/ in *kallas*. The word stress category is expected to be perceived even though the vowel would erroneously be lengthened, since no quantity opposition is involved in the study, but the realizations of these words can add to the finding in Thorén (2006), where there was a tendency by the L2-speaker to lengthen vowels rather than consonants, as was also discussed in 1.5.3.3.

Since speakers in the present study have many different L1's, and some of them do not utilize duration features, it can be expected that some speakers realize syllable prominence by means of tonal gestures and/or intensity difference, the latter of which is known to be of minor importance as phonetic correlate to stress in Swedish. Measurements of relative intensity would also be compromised by different intrinsic intensities in different vowel phonemes (Möbius 2003, Lehiste and Peterson 1959), but will be utilized in the present study if duration and/or tone cannot account for deviant cases. Tonal gestures often accompany stress in Swedish, an issue explored in 1.5.2. L2-speakers of Swedish could however give an indication of how Swedish word stress might be realized in an unambiguous way, using other phonetic correlates, or other combinations of correlates. Research questions in this study:

- Can L2-speakers produce unambiguous realizations of Swedish word stress categories, that don't follow the temporal patterns of native speakers of Swedish? Or, will there be unsuccessful realizations that do follow the temporal patterns of native Swedish speakers?
- Which temporal measure will agree best with native Swedish listeners' categorization into trochaic or iambic categories?
- Do successful L2 realizations utilize the duration of the postvocalic consonant to lengthen stressed syllables in /VC:/ words, to the same extent as native Swedish speakers?

### 3.4.2. Method

#### 3.4.2.1. Material

25 speakers of Swedish as L2 were recorded when pronouncing the Swedish words *racket* [rák:ət] 'racket', *japan* [jəpó:n] 'Japanese', *kallas* [kàl:əs] 'is

called', *kanon* [kənú:n] 'gun/cannon', *person* [pæʂú:n] 'person', *filen* [fi:lən] 'the file', *greven* [grè:vən] 'the count/earl', *banan* [bənón] 'banana'. All of these words are each one member of a minimal word-stress pair in Swedish. Half of the test words were trochaic, i.e. had word stress on the first syllable, and half of them iambic, i.e. had word stress on the second/last syllable. The test words were read from a paper and pronounced in the carrier phrase: *Det var ..... jag menade*, 'It was .... that I meant'. The speakers were also shown pictures or clear examples, to ensure they knew the meaning of the intended word before reading it aloud, but were given no guidance as to the pronunciation (other than the spelling).

The technical specifications concerning recording equipment and signal analysis are identical to those in study 3, previous section. The actual number of recordings for each test word was often 22-23, due to flaws in the technical quality.

### 3.4.2.2. Subjects

The 25 L2-speakers of Swedish were between 17 and 57 years of age, and had 16 different L1's. Their educational levels ranged from less than one year in school to university professor. They had started learning Swedish at ages ranging from 7 to 36, which in most cases means the age at which they came to live in Sweden. Length of residence ranges from 9 months to 24 years. All but one of the L2 speakers had started learning Swedish as adults or in their upper teens.

10 native Swedish speakers, 5 male and 5 female were recorded for comparison, under the same technical circumstances. The native Swedish reference speakers were between 16 and 74 years of age, and spoke varieties of Swedish from some 100 km south of Stockholm, to Kiruna in the very north. None of them had what could be called a strong dialect or accent.

### 3.4.2.3. Native Swedish judges

The recordings of the L2-speakers (in total 182 utterances) were presented in random order to 10 other native Swedes, who marked on an answering sheet which test word they perceived. The response alternatives were 1) the intended word (as 1:st or 2:nd alternative), 2) the other part of the minimal word stress pair (as 1:st or 2:nd alternative), and 3) 'undeterminable'. The native Swedish judges did not know which word of the minimal pair that was the intended word.

### 3.4.2.4. Measures

The recordings were analyzed in Praat (Boersma & Weenink 2001). Durations of vowels and consonants in both syllables (except the word initial consonant), words and utterances were measured. Since all test words ended in a consonant, the spectrograms showed satisfactory boundaries between test word and following ...*jag*... (cf. 2.1). Measurements of f0-maximum, as well as mean intensity over vowels, were carried out in those cases where



unambiguous realizations did not follow an expected temporal pattern, and in versions that followed the L1 temporal pattern, but were not judged as unambiguous by the native Swedish listeners. The word *racket* is omitted from the intensity measurement, since higher intensity in the /a/, compared to /e/ or /ə/ in the second syllable, could possibly be a mere reflection of inherent vowel intensity. In the other test words, the intended stressed syllable contains a vowel of less or same inherent intensity as the unstressed vowel, which means that a higher intensity in the stressed syllable in the latter words is more likely to reflect a perceptually relevant difference.

### 3.4.3. Results

#### 3.4.3.1. Successful realizations

First, a few words about the native Swedish productions. These are regarded as unambiguous without having been presented to a group of listeners, but they are judged as perfectly normal and unambiguous by the present author.

The L2-realizations of the test words, which received 10 responses for “intended word stress category” i.e., all the native Swedish listeners perceived it as the intended word, are those defined as successful. “Unsuccessful” realizations received 0 responses for “intended word stress category”. Recall the discussion in 3.3.3.1 concerning criteria for “successful” and “unsuccessful” realizations. This means that also in this study, a number of ambiguous, but possibly “acceptable” realizations are omitted from the results.

The outcome in terms of native Swedish productions, successful L2-productions and unsuccessful L2-productions, are shown in the figures 3.4.1 and 3.4.2 below. Figure 3.4.1 (left panel) illustrates the results in terms of the intended measure of syllable duration, namely the duration of the VC-sequence in the intended stressed syllable, divided by the duration of the VC-sequence in the intended unstressed syllable (VC/vc). Figure 3.4.1 (right panel) shows the result in terms of the measure that was best correlated to the number of responses for intended word. The latter measure is duration of stressed vowel divided by duration of unstressed vowel (V/v)

Figure 3.4.1 shows great similarities in relative durations for native Swedish speakers and successful L2-productions, and the unsuccessful productions are clearly lower. The value that seems to divide successful productions from unsuccessful ones in both panels is roughly 1, which is the point where the stressed and the unstressed units have the same duration. In both panels a vast majority of successful productions are clearly over 1, indicating that both stressed VC-sequences and vowels are pronounced with longer duration than unstressed ones. There is, however, overlapping in both panels. The two words containing phonologically short vowels in the stressed syllables are included in the figure, and they are also kept apart (successful from unsuccessful) by the V/v-ratio, by roughly 1 for *racket* and 0.7 for *kallas*.

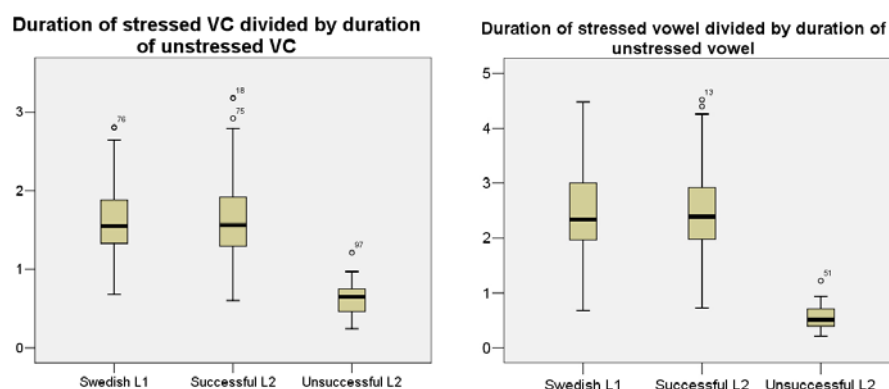


Figure 3.4.1. Left panel: Syllable duration ratios VC/vc for native Swedish productions, successful L2-productions (n= 99), and unsuccessful productions (n= 32). Right panel: Vowel duration ratios V/v, for native Swedish productions, successful L2-productions, and unsuccessful productions.

Figure 3.4.2 illustrates how the relative duration of postvocalic consonants in stressed syllables is distributed over native Swedish, successful, and unsuccessful L2-productions, for the two /VC:/-words *kallas* and *racket*. This measure does not keep the successful, productions separated from the unsuccessful ones, but it does show a clear tendency for successful productions to have longer relative consonant durations than unsuccessful ones. Correlations and significance are reported in table 3.4.1.

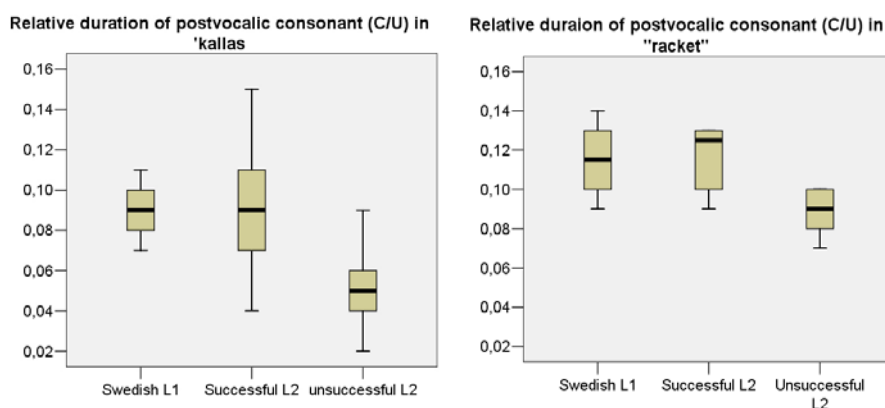


Figure 3.4.2. Left panel: Relative duration of postvocalic consonant in the word *kallas* (n = 10 for all three groups) Right panel: Relative duration of postvocalic consonant in the word *racket*, for native Swedish speakers (n = 10), successful L2-productions (n = 6) and unsuccessful L2-productions (n = 8).

Table 3.4.1 shows how the different relative measures are correlated to “success”, as defined by the number of responses for “intended word stress cate-

gory”. Out of the measures tested on the material, it seems as though the duration of the stressed vowel divided by the duration of the unstressed vowel (V/v) is the most reliable gauge in this material. It was the only one able to separate successful realizations from unsuccessful (for each word), and it was the measure that showed the highest correlation with number of positive responses over the whole material. The second highest correlation is held by vowel duration divided by utterance duration (V/U), implying that mere vowel durations carry more perceptual load as word-stress cue, than the VC-unit. We note however that relative consonant durations are positively – though lower than vowel durations – correlated to number of responses for intended word stress category.

Table 3.4.1. Correlations between number of responses “intended word” and different temporal measures: Duration of VC-sequence in intended stressed syllable, divided by duration of the VC-sequence in intended unstressed syllable (VC/vc), duration the VC-sequence divided by word duration (V/W), duration of the VC-sequence divided by utterance duration (VC/U), duration of the stressed vowel divided by the duration of the unstressed vowel /V/v), by word duration (V/W), by utterance (V/U). For *kallas* and *racket*, duration of consonant following stressed vowel divided by the duration of the word (C/W) and of the whole utterance (C/U).

	VC/vc	VC/W	VC/U	V/v	V/W	V/U	C/W only <i>kallas</i> and <i>racket</i>	C/U only <i>kallas</i> and <i>racket</i>
Correlation with number of responses “intended word”	0.63	0.67	0.6	0.71	0.68	0.7	0.41	0.45
Significance	p < 0.01	p < 0.01	p < 0.01	p < 0.01	p < 0.01	p < 0.01	p < 0.01	p < 0.01

### 3.4.3.2. L2- realizations that do not follow the L1 temporal pattern

There are a few unambiguous realizations by L2-speakers that have durational values (V/v) substantially lower than the lowest realizations of the native Swedish speakers.

In the case of *ba’nan*, all unambiguous L2-realizations had V/v-ratios higher than the lowest L1-value. The 2 unsuccessful L2-versions had both V/v values below 1.

In the case of *’filen*, 4 L2 speakers had V/v-values lower than the lowest L1-realizations, although the mentioned L2-versions were between 1.64 and 1.77, i.e. the stressed vowel had a duration of more than 160% of the unstressed vowel.

In the case of *’greven*, 3 L2 speakers had V/v-values lower than minimum for L1-speakers, although 2 of them had V/v ratios above 1.5, which indi-

cates substantial difference in duration for stressed and unstressed vowel. One L2-speaker had 1.38, and this version is according to the present author unambiguous with respect to word stress, but would have been ambiguous if quantity category had been in focus. This version has an intensity difference of 2.6 dB between first and second vowel, and an f0-maximum in the first syllable (intended carrier of word stress) that is 0.03 bark higher than f0-maximum in the second syllable.

As for *ja'pan*, no successful versions had V/v values lower than L1 minimum, and no unsuccessful versions had values within the range of L1-speakers.

In the case of *'kallas*, the minimum L1-version had V/v as low as 0.68. All successful L2-versions had values above. No unsuccessful L2-version had values within the L1 range for V/v values. Concerning relative consonant duration, being the expected way of lengthening the stressed syllable in /VC:/-words, we see that all successful L2-versions but one, showed C/U-values within the L1 range. The deviant version had a C/U of 0.04 to be compared to the minimum L1 version of 0.07. In the ears of the present author, this version was trochaic, but totally ambiguous with respect to quantity category. V/v and VC/vc values were above L1 minimum for this L2-speaker.

One L2-speaker pronounces the word *ka'non* with VC/vc 1.35 and V/v 1.93, which are values substantially below those of the native Swedish speakers, and is perceived by all Swedish judges as the intended iambic word. The speaker makes a 1.8 dB intensity increase in the intended stressed second syllable, compared to the intended unstressed, first syllable. She also has a mountain-shaped tonal gesture over the second syllable, resulting in a pitch difference of 0.13 bark between max F0 in the first (unstressed) vowel and max F0 in the second (stressed) vowel. The native Swedish speaker with the lowest durational values makes an intensity difference of 2.2 dB between unstressed and stressed vowel plus a distinct tonal gesture LHL (mountain shaped) resulting in a corresponding pitch difference of 0.28 bark between stressed and unstressed vowel. The native Swede with the highest durational values makes an intensity difference of 1.2 dB between unstressed and stressed vowel, plus the (mountain shaped) tonal gesture typical for accent 1 words in focal position (pitch difference 0.53 bark). For the native Swedes in general – for the word *ka'non* – the durational difference seems compulsory, the tonal gesture common, but not always there, and higher intensity (more than 1 dB) in only a couple of cases. The lowest V/v value among L2-realizations does however show nearly double duration for the stressed vowel compared to the unstressed one. No unsuccessful version of *ka'non* had V/v values above 1.

One L2-speaker pronounces the word *per'son* with V/v value (1.31) lower than all native Swedish speakers (min. 1.76), and is still perceived by all Swedish judges as saying the intended iambic word. In this case there is no

clear intensity difference between stressed and unstressed vowel and there is no distinct tonal gesture. The VC/vc ratio is 0.92 that would, according to the hypothesis, have yielded a trochaic response, but the V/v ratio – that seems to correlate better with the perception of native Swedes – was 1.31, which in this case seems to be enough to be perceived as the intended word by all 10 judges. No unsuccessful L2-versions were above V/v 1.

One unambiguous L2-version of *racket* shows a V/v value lower than the lowest value for native Swedish speakers. He does not make a distinct tonal gesture, but makes an intensity difference of 4.6 dB between stressed and unstressed vowel. He also has a C/U ratio of 0.12, which is compared to 0.09, the lowest V/U ratio among the native Swedish reference speakers. He also has a VC/vc value above minimum for L1-versions. Another L2-version of *racket* shows a VC/vc-ratio of 1.21, but has no responses for intended word. The lowest VC/vc-value for Swedish L1 speakers is 1.25. The L2-speaker has however a V/v-value of 0.62, which is less than half the corresponding minimum for L1-speakers. The first VC-sequence is a clear [ak], but the second is only a clear [e] with very unclear traces of a subsequent /t/.

#### 3.4.4. Discussion

The results show that L2-speakers, with few exceptions follow the temporal patterns of native Swedish speakers when they produce unambiguous categories of trochaic and iambic Swedish words. The exceptions seem to consist of deviations from values that determine quantity category rather than word stress category. It seems as though the L2-speakers had a choice with respect to what segment should be lengthened, since the native Swedish judges were not concerned with naturalness or quantity, only stress category. It also seems like native Swedish speakers made a bigger difference in vowel duration, than was required for the signaling of word stress category. L2-speakers could in some cases go below the minimum L1 V/v value, and still have V/v values that safeguarded the word stress category. The great margins in the L1 productions could be explained by the double task of L1 speakers, to signal both word stress category and quantity category. L2-speakers could in this study – figuratively speaking – allow themselves to lengthen vowel and consonant equally, thus increasing the duration of the VC-sequence more than the duration of the vowel. This in turn can explain why the VC/vc ratio was the best gauge in some cases and the V/v in others. This can be exemplified by one L2-version of *racket*, mentioned in the result section. The V/v-value was 1.03, which can be compared to the lowest L1-version of 1.63. 1.03 is very near same duration for stressed and unstressed vowel. It was still judged as the intended trochaic word by all native Swedish listeners. This version was however compensated by a longer consonant duration, yielding a C/U-value of 0.12, compared to L1-minimum 0.09, and a VC/vc-value of 1.48, compared to L1-minimum 1.25

The result points in the direction that duration is a robust correlate to stress, which is in agreement with the claim by Fant & Kruckenberg (1994). This robustness is observed in spite of the fact that there are some factors likely to distort the temporal pattern, e.g. different inherent durations in the segments involved, different quantity categories in words and syllables, and the fact that a few speakers – in spite of the efforts to prevent it – made a small pause after the test word, which is likely to cause final lengthening in the last syllable of the word (Heldner & Strangert 1998). The presence of final lengthening is one plausible explanation to the V/v ratio being more reliable than the VC/vc ratio, since final lengthening in a trochaic word is likely to add more duration to the final consonant than to the second vowel, thus causing more lengthening in unstressed VC-sequences than in mere vowels. Such an increase in duration of an unstressed syllable is not likely to alter the native listener's perception of category, but of course changes measured values.

There is 1 out of 99 unambiguous realizations in this study, where intensity and/or f0-excursions is put forward as a possible but not necessary explanation why an L2-version with “too low” temporal values was still judged as intended word stress category. This implies that word stress in Swedish, has to be signaled temporally, and that dynamic and tonal gestures often accompany the temporal properties.

The observation that a tonal realization often accompanies the word is expected – at least in the native Swedish speakers – since the test words had focal position in the carrier phrase, and since focal position evokes all potential prosodic properties in a word. The impression of the present author is that all L2-speakers pronounced the phrases in a way that made the test word sound as the most prominent word of the phrase, although the realization was not always a typical Swedish focal accent. A greater intensity is sometimes found in the stressed syllable, compared to the unstressed one, which is also expected in focused words, but the opposite case is not rare. According to a standard description of Swedish prosody (Bruce 1977), the test words in the present study would always be in a focal position, which in turn would lead to both a fall and a rise. This could at least be expected in the native Swedish speakers, since they were reading in their L1. The tonal fall was clearly expected in accent 2 words, and possibly in accent 1 words (cf. Engstrand 1995). Many of the L2-speakers were not very familiar with reading aloud, at least not in an L2, and this may have influenced their intuitive feeling for semantic load in the respective words.

The fact that the relative duration of the stressed vowel was the best temporal measure of word stress in this material is not surprising for those words that contain a phonologically long vowel in the stressed syllable, but it was not expected to be the best measure also for the /VC:/-words ‘*kallas* and ‘*racket*, where the main lengthening of the stressed syllable was expected to be found in the postvocalic consonant of the stressed syllable. Duration is in

this material compared for syllables in focal position and unstressed syllables, which should, following e.g. Fant & Kruckenberg (1994), Heldner & Strangert (2001), Bannert (1979a) make the difference in syllable duration even greater. The fact that a consonant has extra duration does not however, necessarily imply that it is the strongest perceptual cue to lexical stress. It seems logical that the vowel carries the main perceptual salience in the case of word stress, provided we assume that the strongest perceptual cue is some kind of syllable duration. 4 out of 8 test words were trochaic, and had consequently a word final postvocalic consonant, following an unstressed vowel. This consonant is likely to influence the VC/vc-value, both by means of inherent duration, e.g. /s/ in *kallas*, (cf. Elert 1964) and as a result of final lengthening in those cases where the speaker makes a pause.

The concluded stronger perceptual weight in vowel duration compared to consonant duration agrees with the findings in the previous study in the present thesis, and with Traunmüller & Bigestans (1988), who found that relative vowel duration is the measure that agrees best with native listeners' categorization into /V:C/ and /VC:/.

The issue of phonetic elasticity is not directly addressed in the present study, but as it is one issue of the present thesis, the absolute durations and standard deviations of phonologically long and short segments are captured in the box plots in figure 3.4.3 below. Standard deviations for phonologically long segments are higher than for phonologically short segments, and the difference in dispersion between long and short segments is bigger for native Swedes than for L2 speakers. This is expected since native Swedes tended to preserve the proper quantity category, while many L2-speakers could produce unambiguous word stress categories with ambiguous quantity categories. The standard deviation among phonologically long segments in the present study is considerably lower than in the previous one (3.3.4).

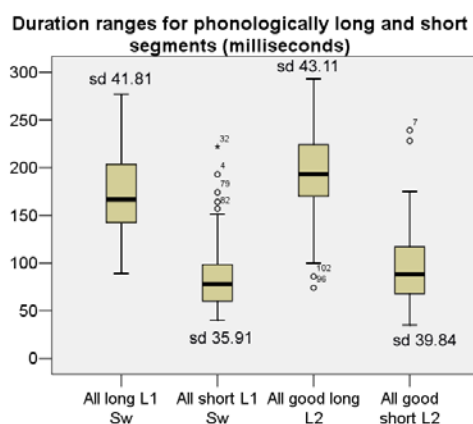


Figure 3.4.3. Duration ranges and standard deviations for phonologically long and short segments, in productions by native Swedish and unambiguous ('good') L2 versions respectively.

Finally we must consider the fact that the Swedish listeners, who had to choose between two possibilities, one trochaic and one iambic, were no judges of how well each realization would function in a real life situation. It was only a decision between three possibilities, 1) trochaic word, 2) iambic word, or 3) “undeterminable”, and the alternatives were written on the answering sheet. We cannot automatically assume that a native Swedish listener who was not prepared for this special task would understand all of the “unambiguous” realizations in the study.

### 3.5. Swedish Accent – Duration of postvocalic consonants in native Swedes speaking English and German

#### 3.5.1. Introduction

The concept of foreign accent – thoroughly discussed in 1.2 – is well known, and in the present study we are interested in temporal properties transferred from the native language to the target language. Recall the quote from Lindblom et al. (1981) in 1.3.1. Prosodic features are claimed to be extra persistent (Gårding 1974b). In studies of L2-speech, the attempt to pronounce an L2, can very well result in temporal values somewhere in between those of the L2-speaker’s L1, and those of native speakers of the target language. This has been shown for VOT in native Spanish speakers’ English (Flege & Efting 1986), and for timing of VC-sequences in Swedish speakers’ L2 productions of English and Japanese (McAllister et al. 2006). According to the feature hypothesis (McAllister et al. 2002), a phonetic property used in a person’s L1 will be easier to learn in a L2 than phonetic properties not utilized by the L1. But what about transfer of properties from L1, which are not needed in the L2? The present study examines the persistency of the Swedish complementary pattern resulting in lengthening of consonants following phonologically short vowels in stressed syllables.

In addition to differences in duration between /V:/ and /V/, Swedish is known to have a complementary consonant duration in stressed syllables (Elert 1964) – to less extent in some southern varieties (see 1.4.3 for a discussion) – resulting in long consonant after short vowel /VC:/ or short consonant after long vowel /V:C/. See 1.7 for an elaborated discussion. If this complementary temporal pattern is analyzed as “compulsory lengthening of stressed syllables”, it can be expected to influence the timing of the VC-sequence in a L2, produced by native Swedish speakers.

English and German have a phonological distinction corresponding to the Swedish quantity distinction, although English is – depending on which author is consulted – often regarded as having a phonemic contrast signaled



mainly by spectral difference between tense and lax vowel, rather than a quantity distinction. The standard varieties of English in UK, USA, Canada and Australia, and German are not known to have a complementary consonant duration after stressed short/lax vowel. Speakers or listeners of the Bavarian dialect, which is reported to have a temporal pattern close to the Swedish (Bannert 1976), did not participate in this study.

It is not obvious which measure of consonant duration, that best corresponds to the native Swede's production, which follows from the discussion in 1.8. The absolute duration of the post-vocalic consonant would be affected by speaking rate and not a reliable measure, and the V/C-ratio, often used in quantity studies, is affected by durational changes in both V and C, or in either V or C, and does not give information on the separate contribution from V or C duration. Segment duration divided by word duration was tested for Swedish (section 3.3), and did correspond to the categorization made by native Swedish listeners, but not as well as vowel duration divided by utterance duration (cf. table 3.3.2). Segment duration divided by utterance duration will be the main measure of relative consonant duration in the present study.

A native speaker of English or German, who is accustomed to Swedish speech prosody and to English or German spoken by native Swedes, can be expected to perceive a stronger Swedish accent as a consequence of more complementary consonant duration, although other phonetic properties in the L2-speech of Swedes are also expected to contribute to the impression of Swedish accent.

The present study puts forward two hypotheses from the reasoning above:

1. Native Swedish speakers will, in their English and German pronunciation, apply a complementary consonant duration that significantly differs from the post-vocalic consonant duration produced by native speakers of Standard English – UK, USA, Canada, Australia – and standard German.
2. Degree of perceived Swedish accent, judged by native users of English and German, will correlate with relative post-vocalic consonant duration.

### 3.5.2. Method

22 native Swedish speakers were recorded when pronouncing the English words *chicken* [tʃɪkən] and *woman* [wʊmən] in the carrier phrase *I said .... again*. 10 native speakers of Standard English (1 Australia, 2 Canada, 3 UK and 4 USA) produced the same utterances. The corresponding procedure was carried out in German, with 18 native Swedes and 10 native speakers of standard German. The words *Mutter* [mutɐ] 'mother' and *kommen* [kʊmən] 'to come', were pronounced in the carrier phrase *Ich habe .... gesagt* 'I said ....' As a control, 12 native Swedes were recorded pronouncing Swedish

words with same or similar VC-sequences as the English and German test words. The Swedish words were: *spiken* [spî:kən] ‘the nail’, *blicken* [blîk:ən] ‘the look’, *boten* [bú:tən] ‘the remedy’, *Otto* [ót:ø] proper name, *omen* [ú:mən] ‘omen’, *lommen* [lóm:ən] ‘the diver/loon’, *pråmen* [pró:mən] ‘the barge’, *kommer* [kóm:ər] ‘is coming’. These words contain the sequences [i:k - ɪk:] to be compared with *chicken*, [u:t - ʊt:] with *Mutter*, [u:m - ʊm:] with *woman*, and [o:m ɔm:] with *kommen*. These Swedish words were pronounced in the carrier phrase *Jag sade .... till dig*. ‘I said .... to you’. The initial sequence *Jag sade* is mostly pronounced as disyllabic [jasɑ]. None of the native Swedish speakers had a south Swedish (Skåne-) dialect (cf. 1.4.3). The recordings were done in sound treated rooms in private homes as well as in the Physiology studio in the University of Stockholm. The recordings were transferred to computer and analyzed in Praat (Boersma & Weenink 2001). Durations of utterances, first vowel in test words as well as post-vocalic consonant were measured. Segmentation of utterances was carried out by means of visual study of waveform and spectrogram, as well as listening. See 2 for further details about segmentation. Stops were measured both including and excluding bursts. If not otherwise explicitly indicated, stop durations in the results refer to occlusion only.

A native speaker of American English and a native speaker of standard German, who had both lived in Sweden for several years, rated the utterances by the Swedish speakers for degree of global Swedish accent, on a scale from 1 “no Swedish accent, or hardly detectable” through 10 “strongest Swedish accent in this sample”. The judges had linguistic education but were not informed about the purpose of the study. The correlation between perceived degree of Swedish accent and relative consonant duration was tested.

### 3.5.3. Results

#### 3.5.3.1. Durations

The native Swedish speakers as a group differed significantly from the native English and German speakers with respect to absolute and relative duration of post-vocalic stops in *chicken* and *Mutter*. For post-vocalic /m/ in *woman* and *kommen*, however, the differences between the speaker groups were smaller, which means that the Swedish speakers lengthened post-vocalic stops more than they did nasals. Total utterance durations were significantly longer for the Swedish L2-speakers than for L1-speakers, and the utterances in German had longer average durations than the English utterances, which is consistent with the fact that the German utterances had one more syllable; 7 compared to 6 for the English utterances. As a group, the native Swedish speakers had longer duration for both V and C in the measured sequences, which is consistent with their longer average utterance dura-

tions, which in turn reflects a slower speaking rate that can be considered natural when speaking an L2.

Figure 3.5.1 illustrates consonant durations divided by utterance durations for the three speaker groups. It shows that speaker groups differ more with respect to stop durations for both English and German (left in panels), than they do with respect to nasal durations (right in the panels). The seemingly shorter stop durations compared to nasal durations produced by L1 speakers are due to exclusion of bursts in the stop duration measure.

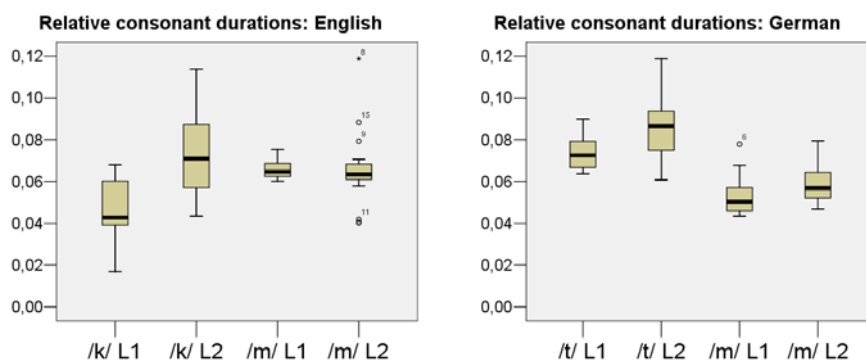


Figure 3.5.1: Consonant durations divided by utterance durations for /k/, /t/ and /m/ for native English speakers (left panel L1), native Swedes speaking English (left panel L2), native German speakers (right panel L1) and native Swedes speaking German (right panel L2).

In figure 3.5.2, vowel and consonant durations for L1 and L2 productions of the test words are shown, and it is clear that the native Swedish speakers pronounce the English and German test words with shorter vowel as well as consonant durations compared to Swedish /VC:/ -words (*blicken*, *Otto*, *lommen*, *kommer*), and it also shows that the Swedish speakers apply substantial complementary consonant lengthening in the Swedish words.

### 3.5.3.2. Correlations

The correlation between “degree of global Swedish accent” rating, and relative consonant duration, defined as occlusion/utterance duration for /k/ and /t/, and total duration of /m/ divided by utterance duration, is significant when pooled, but rather low, and not significant for all consonant types and languages separately. Quite unexpectedly the highest correlations between accent-rating and consonant duration was found for absolute consonant durations. Correlations and p-values are shown in table 3.5.1.

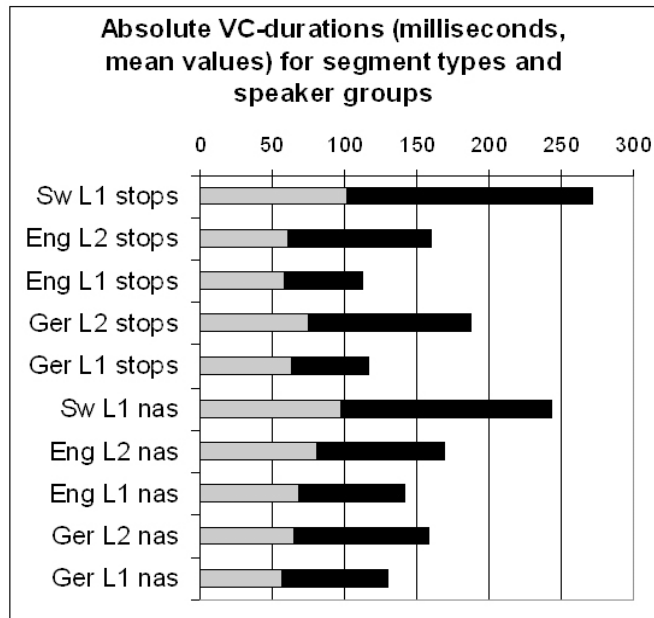


Figure 3.5.2: Absolute vowel and consonant (mean) durations for test words containing stops and nasals after short vowel. Gray portions (left) show vowel durations and black portions (right) show consonant durations.

Table 3.5.1: Correlation coefficients (Pearson) between absolute and relative durations for post-vocalic consonants, and ratings of global Swedish accent. Dur = duration, nas = nasals, u = utterance.

	Stop dur	Nas dur	Stop/u	Nas/u	All C dur	All C/u
English	0.65 p<0.01	0.51 p<0.05	0.48 p<0.05	0.09 p=0.7	0.60 p<0.01	0.33 p<0.05
German	0.37 p=0.14	0.48 p<0.05	0.22 p=0.39	0.27 p=0.28	0.47 p<0.01	0.32 p<0.05

#### 3.5.4. Discussion

The results give some support for the hypothesis that native Swedish speakers would produce complementary longer consonants after perceived short vowel in English and German, than native English and German speakers.

The Swedish complementary pattern was obvious for *chicken* and *Mutter*, showing significantly longer relative durations in the L2-productions, but vague for *kommen* and practically non-existent for *woman*. A possible explanation to this asymmetry could be a general tendency for nasals to have shorter durations than stops. Elert's (1964: 150) data indicate that: "Nasals, such as /m n rn ŋ/, seem to be shorter when compared with homorganic voiced plosives". However, the Swedish words produced by the native Swedish speakers do not support this explanation, since relative nasal duration after a short vowel was well on a par with relative stop durations in the same position. Another explanation to the inconsistent L2-productions by the native Swedish speakers in this study could be that the Swedish speakers perceived the English and German words containing /m/ as having long vowel quantity, and hence would pronounce them without the post-vocalic long consonant. If the Swedish speakers had perceived the words *woman* and *kommen* as carrying phonologically long (or tense) vowel, they would be expected to pronounce the vowel with durations near those of Swedish words with phonologically long vowel, but again, the native Swedes pronounced the vowels in *woman* and *kommen* slightly longer than did native speakers of English and German, but significantly shorter than the phonologically short vowels in the corresponding Swedish words. This indicates that *woman* and *kommen* are indeed perceived by the native Swedish speakers as having phonologically short vowel. It can be assumed that most Swedes are more exposed to native spoken English compared to German, and this state of things could possibly explain more native-like English productions by the Swedish speakers, but it would nevertheless be expected to have the same influence on stops as on nasals. An extended study including fricatives could shed some more light on these issues. Grønnum and Strangert (2007, personal communication) suggest that an L2-learner would easier perceive the (shorter) duration of a sonorant nasal, than of a voiceless stop, whose major duration is realized by silence. This thought seems quite attractive, but again, this might leave even less duration of a "clean" nasal to be measured by the L2-learners perception, which in turn could account for the fact that the shortness of nasals in English and German, was better perceived and imitated by the native Swedish L2-speakers, than the corresponding shortness of voiceless stops.

With regard to correlations between absolute and relative consonant duration and perceived degree of Swedish accent, it is obvious that more factors than post-vocalic consonant duration influence the degree of perceived accent, but the significant correlations indicate that the Swedish accent rating is not independent of the parameter of interest. The result of the ratings also shows that the Swedish speakers who received the lowest ratings for degree of Swedish accent (most native-like) never had relative stop durations exceeding those of the native English and German speakers. Although long consonant duration seems to be a sufficient condition to prevent low scores

on perceived Swedish accent, the opposite is not true, since some of the Swedish speakers received high ratings for Swedish accent, in spite of the fact that they produced postvocalic consonant duration not exceeding those of native English and German speakers. This is expected since many different phonetic factors may contribute to a strong foreign accent. The conclusion is that the Swedish temporal pattern of long consonant after phonologically short vowel is transferred to L2-productions, but not in a uniform way. Swedish speakers of German (and English) may be comforted by the fact that native German listeners are not disturbed by the typical Swedish lengthening of postvocalic consonants (Traunmüller 2007, personal communication). More consonant phonemes should be studied with respect to this phenomenon. A similar study with more, maybe 10 judges of Swedish accent for each language, would show whether the impression of accent as a function of consonant duration is a general phenomenon.

## 4. General discussion and preliminary conclusions

### 4.1. Priority of certain aspects of L2-phonetics over others

In the introduction it was stated that it is impossible to cover all aspects of L2-pronunciation in an educational situation, neither by the teacher, nor by the learner. This is because of restricted time, and the richness of detail there is in the phonology and phonetic realization of any language. Furthermore, the pronunciation teaching shares the time with grammar, vocabulary, conversation etc. The discussion in the present thesis addresses mainly the situation of adult L2-learners, since younger learners in many cases seem to be able to acquire a native-like L2-pronunciation more or less automatically. It was also shown by Flege et al. (1995) and Abrahamsson & Hyldenstam (2006) that the older a person is when starting to learn an L2, the greater the probability of a foreign accented L2-speech, and the accent will not diminish substantially just because the learner spends more time in the L2 environment (Piske et al. 2001). Taken together, this implies that most adult learners of Swedish as an L2 will finish the language course with a foreign accent, containing elements of L1 phonology and/or L1 ways of realizing a common L1-L2 category. Although very little research has been conducted that addresses the impact of different teaching approaches on the learning outcome, some of the work cited in 1.3.2, e.g. Moyer (1999) and Missaglia (1999) indicates that prosodic ingredients in the language instruction often resulted in a better overall pronunciation in L2-learners. This in turn, makes us feel confident that the priorities made in the teaching curriculum has an impact on the learning outcome, i.e. the result is not only dependent on the factors inherent in the learner, and language instruction is not merely a matter of exposing the learners to the target language.

In section 1.2 different aspects of foreign accent were discussed, and from Munro & Derwing (1995) we adopt 3 variables, which can be associated with degree of foreign accent:

- *Global foreign accent*: A listener's perception of how different a speaker's accent is from that of the L1 community.

- *Comprehensibility*: A listener's perception of how difficult it is to understand an utterance.
- *Intelligibility*: The extent to which a listener actually understands an utterance. Tested by % words correct reproduced. Also correct answer to questions, and the time it takes to reproduce or answer.
- With the above variables associated with foreign accent as a point of departure, the present thesis adopts a notion of *listener friendliness* (Bannert 1979c: 132), which pertains to an L2-speech that has high degree of *intelligibility*, but could hypothetically attain any degree of *global foreign accent* (cf. 1.2.4). The present thesis discusses the idea that certain phonetic features, probably not the same for all languages, are crucial for a listener friendly L2-speech. A listener friendly L2-speech is by necessity somewhere between *intelligible* and *native-like*. This location along the accent-ness-variable is perhaps best captured by Abercrombie (1949), who writes "comfortably intelligible". "Comfortably" is interpreted as "not disturbing". An important dividing line is drawn between *naturalness* and *listener friendliness*. Maximal naturalness entails native-like pronunciation, while maximal listener friendliness may or may not do that.

The present thesis suggests a highly simplified strategy for the instruction of pronunciation in Swedish as an L2. The strategy originates from the experience and developing works of a handful of teachers and researchers in the field of Swedish education for immigrants from the 1970's and onwards (Gårding & Bannert 1979, Bannert 1979, 1980, Kjellin 1978, Thorén 1988, Slagbrand & Thorén 1997). The present thesis advocates a further simplification of a couple of existing approaches, and the present one is based on a description of Swedish phonology, that only takes temporal aspects of prosody into account. The present approach is called Basic Prosody (BP), and the present thesis is an attempt to examine BP in the light of existing knowledge and five studies in perception and production.

In addition to sets of vowels and consonants, Swedish has three distinct prosodic contrasts: word stress<sup>11</sup>, quantity and word accent (cf. table 1.1). Ideally the learner would acquire all the segmental and prosodic contrasts and realize them in a "Swedish way", i.e. apply the same phonetic correlates, in the same combinations and in the same weights as native Swedish speakers. The experience of many teachers, and researchers is however, as mentioned above, that adult L2-learners tend to have a foreign accent in their L2-speech. The present author, and others before him, e.g. Bannert (1979c), thought that foreign accented Swedish could be quite acceptable in some cases and quite disturbing in others, and started to listen more closely to find out what was present when the foreign accent was acceptable, and what was absent when the foreign accent was not acceptable. This rather intuitive examination resulted in priority of prosody over segments, with some empirical

---

<sup>11</sup> Stress contrasts on phrase level are also possible, e.g. '*Hälsa på*...' 'greet' *hälsa på* 'visit'.



support with respect to word stress (Bannert 1986, Field 2005). The interest in prosody can be interpreted as a reaction to the earlier tradition of nearly pure segmental interest. The present author drew – from own experience – the conclusion that Swedish speech rhythm, i.e. the temporal organization of spoken Swedish, is crucial to intelligibility, while tonal features would enhance naturalness. During more than 20 years of teaching, the present author made suggestions to improvement of students' pronunciation mainly as suggestions to increase duration in one or more segments. The result of these instructions, manifested as student pronunciation, was so encouraging that a couple of manuals were written, giving highest priority to “stress and length” (Slagbrand & Thorén 1982, 1997, Thorén 1988). The name *Basprosodi* ‘basic prosody’ (BP) was given to the first of these works.

BP results in a teaching approach that amounts to assigning extra length to the proper segments of the speech. By “length” is meant both phonological length and physical duration, and by “proper segments” is meant phonologically long segments in stressed syllables, according to the complementary vowel-consonant length pattern /V:C/-/VC:/ in Swedish (e.g. Aurén 1869, Elert 1964). The learner is presented with rules of a phonologic sort, and at the same time, the aim of the instruction is that the learner produces longer duration in specific segments. The instruction containing the notion of length is assumed to be easily associated with duration by the learner. The extra duration on proper segments is assumed to signal both stress and quantity, and the findings from Bannert (1986) provides empirical support for the assumption that at least word stress is crucial for intelligibility. To assign sufficient duration to a syllable containing a phonologically short vowel, one has to increase the duration of the postvocalic consonant. BP emphasizes the complementary vowel-consonant length pattern, partly because the role of postvocalic consonant duration is not fully acknowledged in the educational field of Swedish as an L2 (Håkansson & Stenquist 1989, Ballardini et al. 1997, 2001). The inclusion of the postvocalic consonant in the distribution of extra duration, priority of temporal aspects over tonal and spectral aspects of Swedish prosody as a foundation for a pronunciation teaching approach, is defended in the introduction of the present thesis, by presenting a number of circumstances. The first category of circumstances is the experience and opinions of some teachers and researchers, who had all tried a prosody oriented teaching of pronunciation to L2-learners of Swedish. Empirical findings from research on English and German reveal that prosody-oriented teaching can yield better results than segment-oriented teaching. Another very important circumstance is the inherent structure of Swedish prosody, which shows greater stability among temporal features, than among tonal and spectral features, seen over the entire Swedish speech community.

## 4.2. Priority of temporal aspects over tonal

The reader is referred to table 1.4, which reflects the higher number of positions where temporal features are applied, compared to tonal. This is one circumstance considered to support the priority. Within the prosodic phonology of a single dialect we recall from 1.5.1, that there are three phonological levels of stressed syllables, depending on information weight and position in sentence and word. The lowest of them, secondary stress, does not require any tonal gesture at all. The intermediate level, accentuated, is according to Bruce (1977) furnished with a tonal fall, although this fall is not generally found in accent 1, e.g. Engstrand (1995). In the highest degree, focal accent, the word accent fall is followed by a focal rise. The phenomenon of stress is always associated with phonological quantity, which is present in all these levels, and entails the presence of a phonologically long segment, regularly manifested as increased duration. Another is that the word accent contrast is not found in all Swedish dialects, and in those dialects where it is found, there are at least 4 main types differing in timing or number of tonal peaks (Gårding & Lindblad 1973). Furthermore the distribution of the word accent categories is not uniform over the Swedish speaking community. The average Swede is thus expected to have heard and understood a great variety of tonal gestures in one and the same word. Therefore an L2-learner is rather likely to produce acceptable tonal gestures by mere chance. There are occasional cases of variation in stress patterns and quantity categories, but nowhere near the variation known among word accent realization and distribution. If we consider the acoustic correlates of stress, we are reminded that increased syllable duration was found to be the most reliable correlate to stress, according to Fant & Kruckenberg (1994). This conclusion was drawn from studies where correlates of perceived prominence was related to temporal, tonal and dynamic correlates, irrespective of phonological stress category. Furthermore, the tonal structure is defined from the temporal structure. Rules for the distribution and realization of accent 1 and 2 must be based on the stress patterns of words, as shown in e.g. Bruce (1977: 18, 50). The claim of the present thesis is not that deviant tonal patterns of Swedish would not bring naturalness down, but that temporal deviations would harm naturalness, intelligibility and, above all, general listener friendliness more.

The arguments in the present thesis concerning stress and word accent, deal with priority among phonological contrasts, and also among their realizations. As discussed in 1.5.3, the levels of phonology and phonetic realizations cannot be expected to be kept apart in the L2 educational setting, and the term “length” is assumed to work on both levels, making it a robust tool for the learner. The present thesis claims that the temporal realization of the quantity contrast is more important to intelligibility and listener friendliness than its spectral realization.

Moreover does the result of the fourth study (section 1.4) indicate that quite a few unambiguous L2-realizations of word stress are made without

any specific tonal gesture, and others with many different tonal gestures, exactly as is the case in L1 Swedish. It is however hard to claim that a sentence is produced without a tonal gesture, since no version in section 6 has a totally stable  $f_0$  throughout the whole utterance. We can only state that some of the unambiguous L2-productions – and L1 productions – had extremely small variations in  $f_0$ , compared to most of the L1 productions, and they also sounded very monotonous. We can also state that a Swedish priest, during a service, can sing a rather long utterance on one single tone, which is assumed to correspond to a stable  $f_0$  frequency. It does not sound natural as speech, but it is intelligible, maybe even listener friendly. And, as mentioned in 1.6.1, one cannot compare a stable  $f_0$  to a hypothetical corresponding monotony of temporal organization. Would in such a case all syllables be of equal duration, or would all segments be of equal duration? If all segments were to be perceived as having the same duration, they would then probably not have it acoustically, since we are used to compensating for different inherent durations, as we perceive the segments of speech. A comparison such as just described, that would reflect a consensus among researchers is highly unlikely if not impossible. This incompatibility raises obstacles when researchers attempt to compare the contributions of duration and  $f_0$  respectively, to intelligibility and listener friendliness.

The study agreed with earlier findings that duration is the most generally reliable correlate to stress, and a number of relative temporal measures showed significant correlation with the perceived stress pattern, ranging from low to moderate. For the whole material in the study, all ratios containing the duration of the VC sequence or the vowel duration – for /VC:/-words also C/W and C/U – showed a significant ( $p < 0.01$ ) correlation with number of “intended word” responses;  $r$  from 0.6 to 0.71 for /V:C/-words and 0.41 and 0.45 for /VC:/-words. To gain more knowledge about stress correlates, a future study should be carried out, with the test word in a non-focal position.

### 4.3. Priority of temporal aspects over spectral

The spectral differences between long and short vowel allophones can be compared to distribution and realizations of the word accent contrast, in that both vary substantially between regional varieties of Swedish. The temporal aspects of stress and the temporal aspects of the quantity contrast are seen as a uniting factor for these regional varieties, whereas tonal and spectral aspects of prosody tend to distinguish these varieties. This issue has been discussed in 1.5.2, with respect to previously known facts, including some “within-system” circumstances, as well as “between dialect” circumstances. This is also of significance when writing e.g. a pronunciation manual for L2-Swedish purposes, which is not only for use in the central parts of Sweden (Stockholm, Uppsala). A distinct priority of phonetic features that agree with

most regional varieties is assumed to be an advantage. If a pronunciation manual describes all segmental and prosodic detail, presented as equally important, the teacher and the students may notice a difference between the pronunciation in the manual, and that spoken in their environment.

First, there are durational differences between /V:C/ and /VC:/ words agreeing with the phonological length (Elert 1964), which can be described as equal length of the whole VC-sequence in both /V:C/ and /VC:/ words, with the difference that the boundary between V and C comes later in /V:C/ words, and earlier in /VC:/ words. This is a simplification with respect to actual durational relations, but assumed as appropriate in an educational context. Furthermore the durational differences between /V/ and /V:/ apply for all nine vowel phonemes. Although there is some variation in V/V: ratio between vowel phonemes, roughly corresponding to spectral differences between /V/ and /V:/ (cf. table 4.1), there is a well-known difference in vowel spectrum or timbre between long and short vowel allophone. When this difference is exemplified, it is typically done with examples containing /a/ or /ʌ/, which are known for the greatest spectral differences between long and short allophone. Common word pairs are *kal* [kɔ:l] ‘bare/bald’ – *kall* [kal:] ‘cold’ and *ful* [fʌ:l] ‘ugly’ – *full* [fəl:] ‘full/drunken’ respectively. The prevalent use of IPA characters reflects the perceivable difference in timbre between the long and short allophone. In the word pair *väg* [vɛ:g] ‘road’ – *vägg* [vɛ:g] ‘wall’, the vowel is transcribed with the same IPA character, reflecting no or minimal spectral difference. The BP approach makes two claims with respect to duration and spectrum pertaining to the Swedish quantity contrast: First, the durational difference (involving both vowel and consonant) between /V:C/ and /VC:/ is a safer commitment in the educational context than the spectral difference, and second, that the Swedish vowel system consists of nine vowel phonemes that combine with a phoneme of length, to yield long counterparts, and not, as suggested by Hadding-Kock & Abramson (1964: 106) “...a subset of eight vowel phonemes that combine with a phoneme of length, to yield long counterparts, and another subset of two vowel phonemes /ə/ and /ʌ/<sup>12</sup>, distinguished by quality.”

So why should the durational realization of quantity be prioritized over spectral? The quantity system shows variation between the vowel phonemes with respect to spectral differences between long and short allophone. This spectral variation can be seen as greater than the corresponding durational variation, although these two variables are not exactly comparable (cf. table 4.1. below). Moreover, as discussed in 1.5.2, there are Swedish dialects that make no or little, or other kinds of spectral differences between long and short allophones: Standard Finland-Swedish uses [a] for both long and short allophone, and [ʌ] for both allophones of /ʌ/. (Reuter 1971). Some southwest Swedish dialects use [ɒ] for both long and short allophone of /a/, and among

<sup>12</sup> Hadding-Kock & Abramson (1964) used other IPA characters to represent the segments, which are not available for the present author.

groups of speakers in the South-Swedish area between Helsingborg and Lund, [ɐ] is used for both long and short allophone (Ingers 1970: 43). This state of affairs implies that durational patterns are more stable as correlates to the quantity distinction than spectral. It is however acknowledged that the South Swedish dialect(s) of Skåne (South Swedish province) shows somewhat different durational patterns compared to central standard Swedish, which manifests itself mainly as a shorter postvocalic consonant duration in /VC:/ words, than the rest of the Swedish speaking community. This distinction between Skåne and the rest of the Swedish varieties is discussed in detail in Gårding et al. (1974), an article with the expressive title “Do the inhabitants of Skåne speak Swedish?” (translation by the present author). Recall however the discussion in 1.4.3, where the findings of Schaeffler (2005) suggested other boundaries between types of quantity realization in the Swedish speaking community. It is admitted that local varieties of Swedish are not in total agreement with BP, but compared to many pronunciation manuals, which account only for a “Central Swedish” or “Stockholm accent”, the BP approach covers considerably more of the Swedish speaking community.

When it comes to the number of Swedish vowel phonemes, the reader is referred to the discussion in 1.5.2, where the chief arguments in favor of a 9 vowel system (vs. a 10 or 18 vowel system), are taken from Eliasson (1985), who points out that long or short vowel allophones are not strictly lexically defined, but rather prove to be stress dependent, as phonologically long segments can be found in different syllables of a lemma, as a result of derivation or compound.

The first experiment in the present thesis (section 3.1) has addressed this issue, and a cautious conclusion of the study, is that probably all Swedish vowel phonemes can be perceived as being either long or short, based on mere temporal properties of the VC-sequence, and irrespective of spectral difference between /V/ and /V:/. Both manipulation experiments (sections 3.1 and 3.2) show that naturalness decreases when temporal manipulations make the listener perceive a long vowel with the timbre of the corresponding short vowel and vice versa. One goal of both experiments was however to test whether it was possible at all, to temporally manipulate words containing the Swedish vowel /ɐ/ and have them perceived as being “wrong quantity category” with respect to timbre. Hadding-Koch & Abramson (1964) concluded that it was impossible, and suggested that /ɐ/ be excluded from the duration governed group of vowel phonemes. The results of the experiment do not support this standpoint, since it revealed that /ɐ/ is also sensitive to duration for quantity categorization. Since the manipulations in this experiment were considerable, it can be argued that going beyond the durations of natural speech – which was done in section 3.1 – is inaccurate, but again, the question was whether it was possible at all, to make native listeners change their categorization of words containing /ɐ/, by means of mere tem-

poral manipulations. And it was. The results of experiment 1 (section 3.1) suggest that the Swedish vowel phonemes line up in a rank order, as shown in 3.1.4, with /ɛ/ being the most dependent on duration for quantity categorization, and the least dependent on spectrum, and /ʉ/ being the least dependent on duration for quantity categorization, and the most dependent on spectrum. The durational differences between long and short vowel allophones, found by Elert (1964) and shown in table 1.4 and in table 4.1 below, show a strong tendency to agree with the rank order in table 1.4. And here, again, we can see a continuum from /ɛ/ to /ʉ/, rather than different kinds of vowels. The quantity categories are accompanied by both spectrum and duration in natural speech, irrespective of which perceptual cue that has the greatest perceptual weight.

Table 4.1. The duration/spectrum dependence of 6 Swedish vowels (the present thesis), compared to V/V: ratios found by Elert (1964), mean values for words in sentences.

Duration/spectrum dependence	/ɛ/	/ø/	/o/	/i/	/a/	/ʉ/
V/V: ratio in % Low ratio means big difference in duration	62	66	69	70	69	77
Vowels ordered according to V/V: ratio	/ɛ/	/ø/	/o/	/i/	/a/	/ʉ/
	62	66	69	69	70	77

#### 4.4. The inclusion of postvocalic consonant “length” in a pedagogic approach

This issue has been addressed more or less in all of the present studies, but we can start with the description of Swedish prosody in the pedagogical setting. A stressed syllable should be pronounced with a longer duration than an unstressed syllable. How is the learner supposed to realize the extra length in the stressed syllables of *vila* [vɪ:lɐ] ‘rest’ and *villa* [vɪl:vɐ] ‘villa/house’ respectively? *Vila* has a phonologically long vowel. Lengthen the vowel. *Villa* has a phonologically long consonant. Lengthen the postvocalic consonant. This is a rule that is likely to make sense to the learner, and is assumed to be easily understood and easily applied, i.e. intuitively realized as longer duration, and also in better agreement with spelling. The standard description for native Swedish children learning to write, is that “double consonant (in writing) means short vowel” – end of story. If BP were to be related to writing, it would naturally relate the double spelling of consonants to the pronunciation of the very segment they represent. And, as found by Elert (1964), and

shown in section 3.5, the complementary length relation /V:C/ - /VC:/, is a persistent temporal production pattern of Swedish<sup>13</sup>. It is thus concluded that the length of the postvocalic consonant – seen both as phonological length and as temporally realized – makes the description of Swedish prosody more logical and consistent, than would be the case without it.

Apart from giving the theoretical description greater stringency, some more possible advantages can be concluded from the studies in the present thesis. Study No 2 (section 3.2) showed that the duration of the postvocalic consonant contributed to the listeners' categorization in /V:C/ and /VC:/, not only when vowel durations were in between typical values for the two quantity categories (figure 3.2.1). This means that when reducing vowel duration in order to make listeners perceive a /VC:/-word, instead of an original /V:C/-word, a smaller adjustment of vowel duration sufficed as the duration of the following consonant increased. This phenomenon appeared also when manipulating the durations from original /VC:/-words in order to have them perceived as /V:C/-words. In the latter case, there was an increase in vowel duration, and the increase in vowel duration could be smaller as consonant duration decreased. Manipulations and the mentioned results are shown in table 3.2.1 and figure 3.2.1. The experiment also showed that "correct" consonant duration contributed to higher degree of naturalness, substantially for /ε/ and moderate for /ʌ/, which were the two vowel phonemes examined in the study.

In the two studies (sections 3.3 and 3.4.), where unambiguous productions of quantity and stress categories by L2-speakers, were compared to those of native Swedish speakers, the successful L2-productions of quantity and word stress categories showed durational patterns very similar to the native Swedish productions, involving both vowel and consonant duration, which can be seen in table 3.3.1 and figure 3.4.2. Study 3 (section 3.3) shows that all unambiguous L2-productions had relative consonant durations equal to or above the minimum produced by native Swedish speakers. The 16 unambiguous L2-realizations of the /VC:/-words *kallas* and *racket*, of study 4 (section 3.4), contained one realization with lower C/U-ratio than the minimum for L1-speakers. This version had a V/v-ratio of 1.33, which is well above the corresponding mean value for L1-speakers. The L2-version mentioned would however not, according to the present author, have been unambiguous with respect to quantity category.

The judgments of naturalness in experiments 1 (section 3.1) and 2 (section 3.2) indicate that a certain relationship between vowel and consonant duration tends to yield higher responses for naturalness. This is particularly clear in figure 3.2.2, upper panels, where versions with "correct" vowel duration and "wrong" consonant duration are compared to version with "correct" durations for both vowel and consonant. The conclusion is that al-

---

<sup>13</sup> We disregard at present the less consistent complementary consonant duration in some dialects in Skåne (South Sweden).

though purely phonemically, the quantity categorization made by native Swedish listeners, can be controlled by mere vowel duration, probably related to the duration of neighboring units, the productions that actually come out as unambiguous, seem to contain the complementary VC duration pattern. The present assumption is also supported by the results of study 5 (section 3.5), which shows that Swedish speakers apply a complementary length pattern when speaking English or German. The pattern is realized as a lengthening of postvocalic consonant after short vowel in stressed syllables. It is not obvious whether this lengthening of postvocalic consonant is a way of confirming the quantity category, or filling out a compulsory duration, proper to the stressed syllable. According to the BP-view, the realization of the quantity distinction can be seen as two alternative ways of distributing the stress-induced extra duration in a syllable.

There are of course many different phonetic features, which could render L2 speech higher naturalness, but in the case of the postvocalic consonant, this is a small cog in the big wheel of a pedagogic prosodic description, and that makes its contribution extra interesting. Recall that the present thesis is supposed to be a major test of the appropriateness of a pedagogic approach, containing a set of priorities and rules. Simplicity and consistency are major guidelines. We also recall from 1.2.4, that BP is designed to promote listener friendliness in the first place, and that listener friendliness is not necessarily equal to naturalness. Total naturalness means native-like, while total listener friendliness can involve more or less perceivable elements of L1 features, with respect both to phonology and to realization.

The reader is once more reminded about the possible interpretation made by practicing teachers, of the rule constructions made by phonologists, which was discussed in 1.5.2. One major effort of phonology, is to account for the largest possible number of phonetic processes, by means of a minimal number of rules, thus defining some linguistic properties as redundant, i.e. automatically following the changes in other features that are defined as independent. The term “automatically” here refers to a programmed condition “if A, then B”, which is not forgotten by a computer or by a set of written rules. It can, however be easily forgotten by a human learner. So if the complementary consonant duration is a “B-phenomenon” and the vowel duration is an “A-phenomenon”, this means of course that also B should happen, although it depends on another event (A). That is also the case with postvocalic consonant duration in the teaching-learning situation; it should be there, since it is, with great generality, there in spoken Swedish (cf. Elert 1964 and study 3.5, this thesis).

From the discussions here and in 4.1.1 it is concluded that the postvocalic consonant is a complementary correlate to both quantity and word stress, provided that the realization of word stress is to signal the intended quantity category as well. Relative consonant duration (C/U) correlated significantly



– though somewhat weaker than vowel duration – with intended quantity category as well as intended word stress category.

#### 4.5. The suitability of BP as basic principle of pronunciation teaching for L2-learners of Swedish

The principles of BP are not conflicting in any crucial way with established descriptions of Swedish prosody, they rather seem derived from them. We recall however from 1.5.3, that there are phonological descriptions and analyses of the postvocalic long consonant after short vowel, that are more or less ignored by the BP-description, since they are not considered to contribute to clearness in the educational situation. Another pedagogic advantage in BP is that the priority of temporal properties does not raise any obstacles with respect to a more elaborate prosodic or segmental description, or learning goal. It merely provides a foundation for it.

The issue of the number of Swedish vowel phonemes is of some interest, considering the task for the L2-learner of Swedish. If we would agree that Swedish has 18 vowel phonemes, they must then consequently have as many descriptions referring to articulations in order to produce correct spectral properties. The spectral properties distinguishing the hypothetical vowel phonemes /v/ and /a/ is likely to be perceivable by the learner (cf. 3.2.1), as would probably /ʉ/ and /ə/, due to their great spectral differences. It would probably prove more difficult for the L2-learner to perceive a spectral difference between a long /ε/ and a short /ε/, which would call for an additional rule to account for a neutralization. The division into 18 vowel phonemes would also require differences in length, in order to render the L2-speech a minimum of naturalness. The rest of the vowel inventory; long /e/ short /ε/, /ø/ , /œ/, /i/, /ɪ/, /o/, /ɔ/, /u/, /ʊ/ would probably have to be described with respect to length in addition to timbre, if the L2-learner should be likely to perceive differences between this great number of phonemes. An inventory of nine vowel phonemes and a prosodeme of length, seems more handy.

An advantage of the BP-description, which includes 9 vowel phonemes, each having a long and a short variety, depending on stress and quantity category, is that some of the spectral differences observed between long and short vowels, are of the kind tense and lax. Spectral differences of this kind can be expected to follow from the speakers' long and short realizations, as a reversed result of the "Extent of Movement Hypothesis" (Fischer-Jørgensen 1964), mentioned in 1.6. Only in this case it is longer duration that allows articulators to move closer to a target. A pedagogical description that takes only durational differences into account, could thus have more or less automatic spectral differences as a consequence. The spectrally oriented educationist could of course argue – still referring to "Extent of Movement Hypothesis" – that having the correct articulatory goals, would automatically

result in durational differences. A central argument in BP is however that giving priority to segment durations (“length” in the educational context) also promotes the signaling of stress, which is a crucial feature with respect to intelligibility (cf. the discussion in 1.3.2 and 1.5.2).

A tempting future study would be to present speech samples from 8-10 L2 speakers of Swedish to e.g. 20 native Swedish judges. The judges should give each sample a rating for comprehensibility and one for general listener friendliness. The ratings could eventually be correlated with measures of relative segments duration in proper syllables of the L2 speech. The speech samples should be chosen to represent an appreciable variation in listener friendliness. A similar possibility could be to present one L2 Swedish speech sample to some 10 native Swedish judges, and a temporally improved version – by means of speech synthesis – of the same speech sample to another native Swedish jury. The ratings could afterwards be compared to find out whether the temporal correction would yield higher ratings for comprehensibility and/or general listener friendliness.

In addition to what is discussed above, there are observations, e.g. as reported by Engstrand (2004:116) indicating shifts in the realization of Swedish front vowels. Young speakers tend to realize /ɛ/ towards /æ/ and /ø/ towards /œ/, which could be taken as another argument in favor of BP, if we define the mentioned shift as yet another instability with respect to spectral properties. In this context it should also be pointed out that temporal properties are not always as stable as BP suggests. A special accent called Swedish on multilingual ground (SMG ‘Svenska på mångspråkig grund’, Lindberg 2006) has developed in suburbs to mainly Stockholm, Gothenburg and Malmö. This accent strikes the present author as applying the same stress pattern as standard Swedish, but realizing stress more or less without lengthening of syllables or segments, making quantity categories unclear, and giving an impression of a rapid and stressful speech. In Bodén & Grosse (2006) it is indicated that many speakers of SMG, also have standard Swedish on their repertoire. It seems likely that there will be a “standard Swedish” for some time ahead, and BP can be a common denominator for most of the regional standard varieties.

To summarize: The present thesis has attempted to evaluate the appropriateness of a pronunciation teaching strategy for Swedish as an L2. Phonological and phonetic circumstances indicate that temporal prosodic features, understood both as phonological contrasts and as their phonetic realizations, are applied in more contexts than the corresponding tonal and spectral features. The empirical studies have shown that all Swedish vowel phonemes can be perceived as non-original quantity category as a result of mere durational manipulations, at least when vowel and postvocalic consonant durations have been manipulated simultaneously in agreement with the complementary pattern /V:C/ - /VC:/. Furthermore, the duration of the postvocalic consonant is found to play a role in the perception of quantity categories.

Two production studies indicated that L2-productions of Swedish quantity and stress had to follow the L1 Swedish temporal pattern in order to be perceived as unambiguous intended category. This pattern includes the duration of the postvocalic consonant with respect to quantity. A finishing production study, where a number of native Swedish speakers pronounced English and German words with complementary lengthening of postvocalic stops, showed that the complementary relation between vowel and consonant duration is a persistent production pattern in Swedish.

# Summary in Swedish (sammanfattning på svenska)

## Inledning

### Bakgrund

Utländsk brytning är ett fenomen som är allom bekant, och som måste ha funnits så länge människor lärt sig andra språk än sitt modersmål. Ibland kan brytningen vara en diskret antydning om att talaren har ett annat modersmål, och ibland kan brytningen göra talet obegripligt eller näst intill. Studier av t.ex. Boyd (2004) har visat att utländsk brytning kan vara en faktor som infödda svenskar kan ha som bedömningsgrund för mer än bara uttalet, och den kan därmed bli ett hinder för yrkeskarriären. I Sverige har undervisningen i svenska som andraspråk uppmärksamats kraftigt sedan 1970-talet, då en strid ström av flyktingar kom in i landet och deltog i svenskundervisning. Uttalet är en del av andraspråksundervisningen, och från slutet av 1970-talet skedde en stark utveckling av metodik och målbeskrivning för uttalsundervisningen. Särskilt fokuserades på svenskans prosodi, dvs. de rytmiska och melodiska sidorna hos språket, även kallat "suprasegmentella drag". Denna avhandlings syfte är att undersöka lämpligheten och korrektheten hos en strategi för uttalsundervisning inom svenska som andraspråk. Undervisningsstrategin bygger på prioritering av temporal prosodi, i detta fall både temporala prosodiska kontraster och de temporala realiseringarna av dessa kontraster. Strategin benämns härnäst "Basprosodi" eller BP. Namnet har tidigare använts av Bannert (1979b), men nuvarande författaren använde benämningen ovetande om detta till ett uttalsläromedel (Slagbrand & Thorén 1982, 1997). BP utvecklades ur behovet av en effektiviserad uttalsundervisning, främst under 1980-talet, som ett försök att förbättra undervisningen för de många politiska flyktingar som tog del av svenskundervisning från tidigt 1970-tal. Ett forskningsprojekt i Lund, under ledning av Robert Bannert och Eva Gårding (Bannert 1979, 1980, Gårding & Bannert 1979), samt en lärobok av Olle Kjellin *Svensk prosodi i praktiken* (Kjellin 1978) har till stor del inspirerat den basprosodimodell som presenteras här.

Intresset för prosodi i invandrarundervisningen i svenska uppstod till stor del på intuitiva grunder utifrån lärares och forskares erfarenhet, men senare studier har gett stöd åt tanken att betoningen är viktig för begripligheten. En studie av Bannert (1986) visar att frasen *I samhället* var mycket svårförstådd för infödda svenskar, när den uttalades med betoning endast på *-häll-*, vilket visar att betoningen är viktig för begripligheten, men det visar inte vilken akustisk egenskap som är huvudansvarig för att signalera betoningen. En liknande studie för engelska gjordes av Field (2005) där felplicerad ordbetoning orsakade försämrade förståelse hos både infödda engelska lyssnare och andraspråksanvändare. Den undersökningen går heller inte in på värdering av olika korrelat till betoningen, men gör en poäng av att inte betrakta prosodin som en enda uttalsvariabel, utan går in på dess olika fonologiska delar. Holm (2006) korrigerade norska yttranden, uttalade av andraspråkstalare. Korrigeringen skedde digitalt, dels endast temporalt, dels endast tonalt, och dels temporalt och tonalt i kombination. Korrigeringen bedömdes av infödda norska lyssnare. Kombinationen av temporal och tonal korrigering bedömdes alltid bäst, men när temporala och tonala korrigeringar skulle bedömas var för sig, bedömdes ibland den tonala och ibland den temporala korrigeringen som bäst, beroende på talarnas förstaspråk. Ett liknande försök gjordes av Bannert (1995) där svenska med utländsk brytning korrigerades temporalt respektive tonalt. Generellt gav tonal korrigering större effekt än temporal, men även här gav temporal och tonal korrigering i kombination bäst resultat. Hos Bannert (1995) finns dock temporala korrigeringar som inte är gjorda helt efter basprosodiska principer; alla fonologiskt långa vokaler är förlängda i den temporalt korrigerade versionen, inga fonologiskt långa konsonanter är förlängda, och i ett fall är en fonologiskt kort vokal förlängd. Ytterligare studier som visar goda resultat av prosodiorienterad undervisning eller vikten av korrekt betoning för begripligheten är Tajima et al. (1997) Hahn (2004) Moyer (1999) Missaglia (1999) och Munro (1995).

En betonad stavelse i svenska kan alltså ha det temporala mönstret /V:C/ som i *kall* eller /VC:/ som i *kall*.

BP prioriterar svenskans ordbetoning och kvantitet, med samma inkludering av postvokalisk konsonantlängd som Kjellin. Att inkludera den komplementära konsonantlängden bland prioriterade uttalsmål antas i föreliggande avhandling utgöra en handfast metod att förlänga betonad stavelse av /VC:-typ. BP tonar i likhet med Bannert ner betydelsen av tonala prosodiska kontraster och deras exakta realisering, främst med argumentet att de inte tycks ha någon avgörande betydelse för begripligheten, samt att både förekomsten och realiseringen av svenskans ordaccenter växlar betydligt mer mellan regionala varianter, än vad de temporala dragen gör.

En annan del av bakgrunden är det faktum att vuxna inlärare sällan lyckas tillägna sig ett andraspråksuttal som liknar infödda talares (Flege m.fl. 1995, Piske m.fl. 2001, och Abrahamsson & Hyltenstam 2006). Då man ibland tvingas acceptera ett inlärningsresultat som inte helt liknar ett infött uttal,

finns det anledning att prioritera uttalsdrag som bidrar till ett lyssnarvänligt uttal. De aktuella undervisningsstrategierna bygger till stor del på positiva erfarenheter när undervisningen fokuserat på svenskans prosodi. Det empiriska stödet för prosodiorienterad undervisning var minimalt då de prosodiska undervisningsstrategierna började ta form, men ovan har redogjorts för några studier som ger stöd för prosodiorienterad uttalsundervisning med engelska, tyska, norska och svenska som målspråk.

Baserat främst på Munro & Derwing (1995) redovisas olika variabler för det ganska vaga begreppet ”utländsk brytning”. ”Global utländsk brytning” står för en subjektiv bedömning gjord av lyssnaren avseende hur starkt ett uttal avviker från förstaspråksnormen. ”Förståelighet” (eng. ’comprehensibility’) är lyssnarens subjektiva bedömning av hur lätt eller svårt det är att förstå den utländska brytningen. ”Begriplighet” (eng. ’intelligibility’) är en objektiv mätning av hur mycket lyssnaren faktiskt har förstått av det som sades med utländsk brytning. En väsentlig upptäckt bland dessa variabler är att förekomsten av ”global utländsk brytning” gör att graden av begriplighet minskar något, men ökad grad av ”global utländsk brytning” är inte korrelerad med ytterligare minskad begriplighet.

Resonemanget kring olika aspekter av utländsk brytning är ett försök att motivera en undervisningsstrategi som kan acceptera kvarstående spår av förstaspråkets fonologi. Filosofin kring BP är att ett *lyssnarvänligt* språk mycket väl kan ha kvar spår av talarens modersmål, men att vissa drag måste vara enligt den svenska målspråksnormen för att talet ska bli begripligt och just lyssnarvänligt. Rent erfarenhetsmässigt är de temporala dragen i svenskan själva förutsättningen för begriplighet och lyssnarvänlighet, och föreliggande avhandling är ett försök att utvärdera BP, dels genom jämförelse med befintliga rön, och dels genom några experiment som kan kasta ytterligare ljus på förhållandet mellan svenskans temporal fonologiska kontraster och deras perceptuella ledtrådar.

## Principer för basprosodi

Basprosodin kan sammanfattas i tre uppmaningar till inläraren av svenska som andraspråk:

- Betona rätt ord i satsen
- Betona rätt stavelse i det betonade ordet (...i de betonade orden) genom att göra den längre än omgivande, obetonade stavelser.
- Förläng rätt ljud i betonade stavelser: vokalen eller närmast följande konsonant.

Eftersom BP betraktar ökad duration som det huvudsakliga medlet för signalering av både betoning och kvantitet, skulle uppmaningarna kunna sammanfattas:

- Förläng rätt ljud.

Detta skulle ställa stora krav på inläraren vad beträffar sats- ord- och stavelsestruktur, men uppmaningen fångar in en av BP:s grundidéer, nämligen att längden, förhoppningsvis realiserad som duration, signalerar både betoning och kvantitet, två prosodiska kontraster som antas vara av stor betydelse för begriplighet och lyssnarvänlighet

Dessa regler och deras inbördes ordning och implikationer antyder att det inte är vokallängden som betraktas som en primär prosodisk egenskap, utan att all längd följer av betoning, och måste ligga antingen hos vokal eller hos konsonant i en betonad stavelse (se figur 1.2. och 1.3). Det ska också sägas att den hierarki av betoningsgrader i svenska, med tillhörande fonologiska kontraster och respektive fonetiska korrelat, som främst följer Bruce (1977, 1998) är av avgörande betydelse för det formella utvecklandet av basprosodistrategin, som var mer intuitiv i sin början. Trots att Bruces (1977) arbete främst är en kartläggning av svenskans tonala ordaccenter, bidrar det i hög grad till att skapa ordning även på övriga områden i den svenska prosodin, och sätta in alla egenskaper i ett satsperspektiv.

### Prioritering av temporala fonologiska kontraster framför tonala

Svenskans prosodi är schematiskt framställd i övre delen av tabell 1.1 (inklusive exempel), och innehåller tre huvudsakliga kontraster: betoning, kvantitet och ordaccentkontrasten. Av dessa är betoning och kvantitet prioriterade i BP-konceptet. Den tonala ordaccentkontrasten är inte prioriterad eller specificerad i BP, men det finns inget hinder för att den tas med i undervisningen; BP utgör snarare grunden för ordaccenterna. Ordaccentkontrasten förekommer inte i alla svenska dialekter, t.ex. inte i standardfinlandssvenskan. Infödda svenskar har i regel inga svårigheter med att förstå finlandssvenska. Vidare finns det åtminstone fyra huvudtyper av tonal realisering av ordaccenterna (Gårding & Lindblad 1973), varav vissa skiljs åt av antalet tontoppar i accent 2, och vissa skiljs åt av den temporala strukturen, dvs. när i de aktuella stavelserna som tonfall eller stigningar sätts in. Distributionen av accent 1 och 2 är inte heller likformig bland de varieteter som har dem. Nuvarande författaren har märkt väldigt tydligt hur en mängd gamla kända accent 2-ord bytte till accent 1, när han flyttade från Stockholm till Sundsvall. De tonala ordaccenterna är även neutraliserade i sång. Detta betyder att svenskans tonala prosodiska kontrast på lexikal nivå kan neutraliseras utan att någon nämnvärd försämring av förståelsen uppstår. Vidare gör sig ordaccenterna gällande endast i de två högsta betoningsgraderna ”accentuerat” och ”fokuserat”, medan de är orealiserade i bibetonad och obetonad ställning, så att t.ex. efterleden ”-tomten” i *jultomten* och *granntomten* har i princip samma tonala struktur, som i sin tur är avhängig av hela ordets prosodi och ordets ställning i satsen eller frasen. Tabell 1.3 illustrerar hur olika

prosodiska kontraster manifesteras vid olika betoningsgrader. I obetonad ställning finns i princip endast en sekvens av ljud och alla prosodiska fonologiska kontraster är neutraliserade. En gråzon mellan betnad och obetonad ställning är lexikaliserade fraser och partikelverb, där t.ex. verbet *känna* i frasen *känna efter*, enligt t.ex. Anward & Linell (1976) är bibetonat, medan informella försök med fonetisk redigering av nuvarande författaren visar att verben *känna* i frasen *känna efter*, och *tjäna* i frasen *tjäna pengar* i ledigt tal, är utbytbara mot varandra, utan att en ovetande lyssnare hör skillnad. Om verbet ska betraktas som bibetonat skall även kvantitetskategori signaleras. Om kvantitetskategorin inte signaleras ska ordet anses obetonat (t.ex. Bruce 1977, 1998).

### Prioritering av temporal realisering av betoning och kvantitet framför tonal respektive spektral

Svenskans betoning realiseras som ökad duration hos den betonade stavelsen (t.ex. Fant & Kruckenberg 1994), med en tonal gest i accentuerad och fokuserad ställning (Bruce 1977), samt med mer sluten vokalartikulation (Engstrand 1988). Dessutom har röstkällan ofta ett spektrum med högre intensitet i högre frekvensområden i betonade vokaler jämfört med obetonade (Sluijter and van Heuven 1996, Marasek 1996). BP tar fasta på den temporala realiseringen, vilket också är det mest stabila korrelatet till svenskans betoning, enligt Fant & Kruckenberg (1994). En tanke är att inlärarna behöver uppmärksammas på den temporala realiseringen, medan de spektrala egenskaperna antas följa mer automatiskt, när inläraren väl lägger ökad energi och duration på rätt segment i talet. Om inläraren förlänger rätt segment i betnad stavelse, kan även kvantitetskategorin realiseras, och en ”fonetisk åtgärd” bidrar till signaleringen av två viktiga fonologiska kontraster.

När det gäller realiseringen av svenskans kvantitetsdistinktion prioriterar BP duration över klangfärg (vokalspektrum). Även här tycks det som att den temporala manifestationen av kvantiteten är mer stabil, dels inom vokalsystemet, och dels bland de regionala varieteterna av svenska. Det ska också påpekas att BP utgår ifrån ett 9-vokalsystem med ett längdprosodem som grund för kvantiteten. Termen ”längd” i pedagogiska sammanhang är tänkt att lätt associeras med duration av inläraren, trots att vi vet att den fonologiska längden (kvantiteten) i svenska uttrycks med hjälp av både duration och klangfärg (vokalspektrum). Vad gäller skillnader inom vokalsystemet, har vokalfonemen /a/ och /ʌ/ välkända och betydande spektrala skillnader i den centrala standardsvenskan, medan /ɛ/ har så gott som samma klangfärg i både lång och kort allofon. Vidare växlar dessa spektrala mönster ganska mycket mellan regionala varieteter av svenska. Ett tydligt exempel på små eller inga spektrala skillnader mellan lång och kort vokalallofon är finlandssvenska, åtminstone den standardvariant som talas i Helsingfors, och som också är väl undersökt (t.ex. Reuter 1971, 1982). I vissa västsvenska dialek-



ter kan mycket väl ett kort /a/ uttalas ”mörkt” dvs. som [ɒ] (egna observationer), och i västra Skåne, från Helsingborg till Lund, kan lång och kort /a/ ha samma kvalitet, dvs. [a] (Ingers 1970). Motsvarande argument som vid realisering av betoningen gäller även här, att talaren genom att realisera kvantiteten främst som ökad duration hos rätt segment, även signalerar betoning. Om förlängning av postvokalisk konsonant efter kort vokal har samma prioritet som vokallängd, finns det i regelverket inbyggt ett robust sätt att förlänga alla betonade stavelser.

## Empiriska studier

De empiriska studierna i föreliggande avhandling har fokus på de perceptuella ledtrådarna till svenskans betoning och kvantitetskontrast. Enligt Fleges (1995) ”Speech Learning Model” (SLM / ”Talinlärningsmodell”) kan en fonetisk kategori (vanligtvis samma som fonem) av en andraspråksinlärare uppfattas med hjälp av andra perceptuella ledtrådar än de som förstaspråkstalaren nyttjar. Detta antas leda till en annan typ av inre representation av den aktuella kategorin, och denna avvikande representation kan leda till ett uttal som avviker från de infödda talarnas. På samma sätt kan en fonologisk kontrast uppfattas, men med hjälp av andra perceptuella ledtrådar som grund för diskrimineringen. Flege (1995) menar vidare att även om representationen är identisk med infödda talares, kan den ändå resultera i avvikande produktion på grund av ovana motoriska mönster, t.ex. hur många konsonanter som kan uppträda ord- eller stavelseinitialt eller –finalt i förstaspråket respektive målspråket. För svenskans del är det intressant att veta om andraspråksinlärare kan realisera betoning och kvantitet på andra sätt än med de typiskt svenska temporala mönstren, och ändå bli väl förstådda. Eftersom det pedagogiska värdet av den postvokaliska konsonantlängden har varit omstritt, riktas särskild uppmärksamhet mot den i föreliggande avhandling görs försök att besvara följande frågor:

1. Är alla vokalfonem känsliga för duration när det gäller perception av kvantitetskategori, eller är det som Hadding-Koch & Abramson (1964) föreslår, att /a/-fonemet endast är känsligt för spektrala egenskaper som perceptuell ledtråd till kvantitetskategorisering, medan övriga vokalfonem realiserar sin kvantitetskategori främst genom duration?
2. Har den postvokaliska konsonantens duration någon inverkan på perception av kvantitetskategori? Hadding-Koch & Abramson (1964) och Behne m. fl (1998), har dragit slutsatsen att manipulation av endast konsonantdurationen inte kan få lyssnare att ändra sin perception av kvantitetskategori. Om konsonantdurationen skulle visa sig ha en roll som perceptuell ledtråd, vore detta ett starkt argument för BP.

3. Kan en andraspråkstalare göra en otvetydig realisering av kvantitetskategorier med andra fonetiska medel än infödda talare av svenska, t.ex. med endast spektrala medel eller med endast vokalduration?
4. Kan en andraspråkstalare göra en otvetydig realisering av ordbetoningskategorier med andra fonetiska medel än infödda talare av svenska, t.ex. med endast tonala, spektrala eller dynamiska medel?
5. Är svenskans komplementära längdmönster så motsståndskraftigt att det förs över till svenskars uttal av andra språk? Observationer av Lindblom (1981) och Engstrand (2007) tyder på det.

### Duration eller spektrum som korrelat till kvantitet

Vilket av kvantitetens fonetiska korrelat är mest pålitligt? Tidigare studier som Hadding-Kock & Abramson (1964) och Behne m. fl. (1997) är ense om att durationen är det generellt pålitligaste korrelatet till kvantitet, men att signaleringen av kvantitetskategori för /a/ och /ʌ/ sker genom ett intrikat samspel mellan duration och spektrum. Hadding-Kock & Abramson (1964) föreslog till och med att /ʌ/ skulle undantas från vokalsystemet med långa och korta allofoner, och utgöra två separata fonem /ʌ/ och /ø/, som åtskildes av endast vokalspektrum.

En stockholmsk kvinna spelades in när hon läste ett antal ord från en lista. Orden utgjorde minimala kvantitetspar, *hut* – *hutt*, *mat* – *matt*, *söt* – *sött*, *tät* – *tätt*, *vit* – *vitt*, *våt* – *vått*. Ordens vokal och finala konsonant manipulerades digitalt i Praat (Boersma & Weenink 2001) så att vokalen i /V:C/-ord fick successivt kortare duration, medan konsonanten fick successivt längre duration. /VC:/-ord manipulerades på omvänt sätt, så att de temporalt kom att likna /V:C/-ord alltmer. 20 infödda svenska lyssnare fick för varje version avgöra om de uppfattade ett /V:C/-ord eller ett /VC:/-ord. För alla undersökta vokalfonem gällde att båda kvantitetskategorierna (/V:C/ och /VC:/) kunde – med enbart durationsförändring – fås att uppfattas som motsatt kvantitetskategori. Dock krävdes mer durationsförändring för /ʌ/ plus efterföljande konsonant, än för övriga vokalfonem. De undersökta vokalfonemen fördelade sig i försöket enligt följande, med avseende på hur mycket durationsförändring som krävdes för att lyssnarna skulle uppfatta dem som icke-ursprunglig kvantitetskategori; minst för /ɛ/ och /ø/, och mest för /ʌ/: /ɛ/ /ø/ << /o/ << /i/ /a/ << /ʌ/. Det betyder att ju längre till höger ett vokalfonem står, desto större roll spelar skillnaden i vokalspektrum mellan lång och kort allofon, för att lyssnaren ska uppfatta kvantitetskategori; /V:C/-ord eller /VC:/. Enligt lyssnarna blev uttalets naturlighet lidande av manipulationerna, men försöket visade ändå att ord innehållande samtliga svenska vokalfonemen, kan manipuleras temporalt så att infödda svenskspråkiga lyssnare uppfattar ursprunglig lång vokal som kort och vice versa. Således kunde även ursprungligt *hut* [hʌ:t] komma att uppfattas som *hutt* [hø:t], trots att /ʌ/-kvalitén var oförändrad, om bara /ʌ:/et var tillräckligt kort och, troligen,

/t/:et tillräckligt långt. På motsvarande sätt uppfattades ursprungligt *hutt* [høt:] som *hut*, om [ø]-ljudet var tillräckligt långt och, troligen, /t/:et tillräckligt kort. Detta antogs av Hadding-Koch & Abramson (1964) vara omöjligt för /u/, men det aktuella experimentet visar att man med enbart durationsförändring kan få ord med alla i experimentet inkluderade vokalfonem att uppfattas som icke ursprunglig kvantitetskategori.

### Den postvokaliska konsonantens duration som en kompletterande perceptuell ledtråd till kvantitetsdistinktionen

Detta försök liknar det förra, men här förändrades orden i två serier. Orden *skuta* – *skutta* och *mäta* – *mätta* fick sina vokaldurationer (endast första vokalen) förändrade stegvis för att få en varaktighet som stämde överens med vokallängden i icke-ursprunglig kvantitetskategori, dvs. det andra ordet i det minimala paret. Ena serien fick endast vokaldurationen förändrad, medan den andra serien fick både vokal- och konsonantduration ändrade, för att temporalt efterlikna icke-ursprunglig kvantitetskategori (se tabell 3.2.1). 30 infödda svenska lyssnare bedömde vilken kvantitetskategori de tyckte sig höra hos vart och ett av stimuli. Resultatet visar att den postvokaliska konsonantens duration har en effekt på lyssnarnas kategorisering, tydligast för de stimuli som hade vokaldurationer i gränslandet mellan /V:/ och /V/ (se figur 3.2.1). Av t.ex. Behne m. fl. (1998: 2294) har föreslagits att den kompletterande konsonantlängden är ”temporal artifacts of the preceding vowel quantity”, vilket kan tolkas som att konsonantdurationens roll är att endast ge den aktuella stavelsen rätt totalduration, så att den inte blir överlång \*(V:C:) eller kort \*(VC). Det aktuella försöket visar att durationen hos den postvokaliska konsonanten är en av alla de ledtrådar som lyssnaren använder sig av för att avgöra kvantitetskategori.

### Andraspråkstälares strategier för realisering av svenskans kvantitetsdistinktion

Denna studie innehåller inspelningar från ca 30 andraspråkstalare av svenska, samt 10 infödda svenska talare. De läste orden: *gratis* – *matta*, *mäta* – *etta*, *vila* – *sillen*, således parvis samma VC-sekvens i betonad stavelse, men med motsatt kvantitetskategori, detta för att talarna inte skulle bli nervösa av att känna igen en bekant svårighet, om de fick se äkta minimala par. Orden uttalades i bärfrasen *Det var ... jag menade*. Talarna hade 18 olika förstaspråk, varav 6 hade någon typ av fonologisk kvantitetskontrast. 10 infödda svenska lyssnare bedömde vilket av två möjliga ord de hörde, utan att veta vilket som var det avsedda ordet. Det fanns även ett svarsalternativ om lyssnaren inte kunde avgöra kategori. Durationer hos bärfras, ord, vokal och konsonant mättes och olika relativa mått beräknades. Resultatet kan sammanfattas så, att de realiseringar som av samtliga 10 infödda svenska lyssna-

re uppfattades som ”avsedd kvantitetskategori”, alltid hade durationer som överensstämde med det infödda svenska mönstret hos 10 svenska kontrolltalare. Detta gäller även durationen hos den postvokaliska konsonanten. De flesta andraspråkstalare som producerade otvetydiga kvantitetskategorier gjorde för /a/ även en spektral skillnad mellan lång och kort allofon, men i genomsnitt tillämpade andraspråkstalarna större temporala skillnader än de infödda svenskarna, och spektrala skillnader i lägre grad och med större spridning i vokalrymden (figur 3.3.4). Detta förhållande är i överensstämmelse med en synpunkt i Strange (1995: 79): ”It may be that temporally cued contrasts are easier to perceive than spectrally cued contrasts.” Lite senare i samma artikel står det dock att “...perception and production mastery may be uncorrelated in more experienced L2 learners.” Dessa synpunkter har relevans även för nästa försök som presenteras nedan.

### Andraspråkstaляres strategier för realisering av svenskans betoningsdistinktion

Denna studie innehåller inspelningar från ca 20 andraspråkstalare av svenska, samt 10 infödda svenska talare. De läste 8 ord som vart och ett har en motsvarighet med betoning på annan stavelse: *ba'nan 'filen 'greven ja'pan 'kallas ka'non per'son* och *'racket*. Orden uttalades i samma bärfras som i föregående studie. 10 infödda svenska lyssnare avgjorde vilket av två möjliga ord de uppfattade, utan att veta vilket ord som var det avsedda, eller svarade att de inte kunde avgöra kategori. Eftersom man kan räkna med en stor variation i realiserandet av orden – talarna var på en mängd olika inlärningsnivåer och hade 15 olika förstaspråk – skulle försöket kunna visa om det går att realisera svenskans betoningskategorier otvetydigt, utan att följa det typiskt infödda svenska sättet att kombinera duration, spektrum och ton. Testorden hade fokusposition i bärfrasen. En postvokalisk konsonant kan ju tillhöra antingen samma stavelse som vokalen eller en följande stavelse beroende på kvantitetskategori. En fonologiskt lång postvokalisk konsonant kan definieras så att dess första del hör till första stavelsen och dess andra del hör till följande stavelse, enligt principen för geminering. En sådan definition skulle vara mycket svår att omsätta i praktiska mätningar; var ska man dra stavelsegränsen inne i ett segment? I enlighet med Fant & Kruckenberg (1989) valdes VC-sekvensen, varvid hela konsonantdurationen ansågs höra till samma stavelse som föregående vokal. I försöket testades om den ”förlängningsbara” VC-sekvensen i respektive betonad stavelse, hade större duration i förhållande till motsvarande ljudsekvens i den obetonade stavelsen i samma ord. Resultatet visar att det finns ganska stark korrelation mellan kvoten betonad VC-sekvens/obetonad VC-sekvens, och antal svenska lyssnare som tyckte sig höra avsett ord. Det var dock en viss överlappning mellan värdena för nämnda kvot, mellan trokéiska (tvåstaviga med betoning på första stavelsen, t.ex. *'racket*) och jambiska ord (tvåstaviga med betoning på

andra stavelsen, t.ex. *ra'ket*), vilket kan förklaras av faktorer som segmentens inherent längd, och final förlängning. Kvoten mellan vokal i betonad stavelse och vokal i obetonad stavelse visade sig vara starkare korrelerad med de infödda lyssnarnas kategorisering (se tabell 3.4.1). Språkljud förlängs ofta före en paus, och många av talarna, även de infödda svenska, gjorde ibland en liten paus efter testordet, innan de fortsatte att uttala bärfrasen. Detta innebär att andra faktorer än betoningen påverkar VC-sekvensens duration.

### Svensk brytning – den postvokaliska konsonantens duration hos svenskar som talar engelska och tyska

Utifrån den typiska temporala strukturen i en svensk betonad stavelse, /V:C/ respektive /VC:/, antogs att svenska talare skulle ge postvokalisk konsonant större relativ duration vid uttal av engelska och tyska ord som upplevdes ha kort vokal, jämfört med infödda talare av engelska och tyska. Både engelska och tyska har en motsvarighet till den svenska kvantitetsdistinktionen, men saknar i de flesta varieteter den komplementära konsonantlängden. Observationer av svenskers sätt att tillämpa komplementär konsonantlängd i engelska och tyska har gjorts av Lindblom m.fl. (1981: 3) och Engstrand (2007: 77).

22 infödda svenskar fick i en bärfras uttala de engelska orden *chicken* och *woman*, och 18 infödda svenskar uttalade de tyska orden *Mutter* och *kommen*. Testorden antogs på intuitiva grunder uppfattas av svenskar som ord med kort vokal, vilket bekräftades av hur de engelska och tyska orden realiserades i förhållande till svenska ord med samma ljudsekvenser. 10 infödda talare av vardera engelska (brittisk, amerikansk, kanadensisk och australisk varietet) och tyska (Tyskland utom Bayern) spelades in som kontrollgrupp. I orden *chicken* och *Mutter*, som innehåller klusil skilde sig svenskarna tydligt från de infödda talarna av engelska och tyska, genom att ge /k/ och /t/ signifikant större relativ duration än hos förstaspråkstalarna. Mätningen skedde i den tysta tillstängda fasen (oklusionsfasen) som föregår /k/- och /t/-explosionen. För de båda orden som innehåller ett postvokaliskt /m/ var bilden mindre tydlig, nära en signifikant skillnad mellan första- och andraspråkstalare i *kommen*, men helt osignifikant för *woman*. En möjlig förklaring till olikheterna för klusil och nasal, skulle kunna vara att de svenskspråkiga talarna trots allt uppfattade *woman* och *kommen* som att de hade lång vokal och därmed kort konsonant. Genom att mäta vokal- och konsonantdurationer hos de svenskspråkiga talarna, i motsvarande svenska ord med samma ljudföljd som i de engelska och tyska orden, står det dock klart att svenskarna producerade alla de engelska och tyska orden som ord med kort vokal.

## Allmänna slutsatser och diskussion

I det som länge varit känt om svenskans prosodi finns indikationer på att de temporala egenskaperna är grundläggande i svensk prosodi; mer grundläggande och generella än ton för betoning, och mer grundläggande och generella än spektrum för kvantitet. Man kan också säga att de temporala egenskaperna förenar de olika svenska regionala varieteterna (dialekterna), medan tonala och spektrala egenskaper specificerar varieteten geografiskt. Det finns ett fåtal undersökningar som säger något om vikten av olika prosodiska egenskaper för att bli förstådd på svenska.

De försök som redovisats i föreliggande avhandling utgör stöd för BP. Första försöket ger stöd för ett 9-vokalsystem med duration som perceptuell ledtråd för kvantitetskategorisering av alla vokalfonem, men med olika interaktion med vokalspektrum för olika vokalfonem. Det andra försöket visar att den postvokaliska konsonantens duration spelar en roll – om än underordnad vokaldurationen – som perceptuell ledtråd till kvantitetskategorisering. Det tredje försöket visade att det sannolikt är nödvändigt att involvera både vokal- och konsonantduration för att signalera kvantitetskategori på ett otvetydigt sätt, när inte kontexten är till någon vägledning. På liknande sätt visade det fjärde försöket att durationsskillnader mellan betonade och obetonade stavelser fanns med i alla första- och andraspråksproduktioner av otvetydig ordbetoningskategori. Dock fanns det fall då andraspråkstalarna signalerade avsedd betoningsskategori, men otydlig eller annan kvantitetskategori, på grund av annan fördelning av duration inom den betonade VC-sekvensen. Det femte försöket visade att det svenska komplementära mönstret /V:C/-/VC:/ är så etablerat i de flesta svenskers talmotorik, att de tillämpar det även i engelska och tyska, där mönstret inte finns i standardvarieteterna. Av de redogjorda strukturella förhållandena och av resultaten från undersökningarna i föreliggande avhandling, dras slutsatsen att basprosodi-strategin är i samklang med hittills kända rön, och att den komplementära konsonantlängden har en dubbel roll, dels som ”buffert” för rätt duration hos betonade stavelser, och dels som ytterligare en temporal ledtråd till perception av kvantitetskategori, utöver vokalens duration. Man kan inte utifrån det som redogjorts här anse det bevisat att temporala drag gör större nytta än spektrala och tonala, som förbättrare av ett andraspråksuttal på svenska, men som undervisande lärare med goda intuitiva erfarenheter av basprosodi-strategin, kan man konstatera att inget talar emot den, och att den innehåller något av en minsta gemensam nämnare för svensk prosodi. Det behövs mer forskning som relaterar den brutna svenskans begriplighet och lyssnarvänlighet till väldefinierade akustiska variabler. Svenskans fonotax, som tillåter tunga konsonantkluster, är troligen också en mycket viktig faktor när det gäller att bli förstådd på svenska, och den finns inte med i en basprosodisk strategi.

# Acknowledgements

I wish to thank my 1:st supervisor Bob McAllister, and Hartmut Traunmüller, my assistant supervisor, who were always available and guided me through the process. With Bob I had many inspiring discussions, and he gave me ideas, a lot of freedom, and had a most delicate way of steering me back up on the track, when I was doing too wide excursions in the prosodic and pedagogic landscape. He has also continuously checked my English, but remaining errors are of course my own responsibility.

If Robert Bannert hadn't helped me getting started with manipulations of foreign accented speech at Umeå University in the 1990's, my doctoral studies would probably not have commenced at all. Eva Strangert kindly guided me through my master paper also in Umeå, and without the technical assistance of Thierry Deschamps, there would not have been much manipulation done.

The experts on phonetics and second language speech learning, Olle Kjellin and Björn Hammarberg gave me lots of ideas and inspiration in the discussions we had. My mother and my very special friend Göran Söderlund gave me shelter in Stockholm, which was invaluable. Furthermore, Göran always challenged my ideas, whatever they were, and encouraged me to continue.

To do most of the work 400 kilometers away from the university had its pros and cons. A good thing was the total peace and quiet as my family was at work or at school most of the time. Thank you also Jan-Åke Olsson for kind and cunning technical support when computers wouldn't cooperate and Peter Branderud, Bosse Kassling and Hassan Djamshidpey – the technical rescue team of the linguistic department at Stockholm University – were too far away – thank you guys for your skill, swiftness and patience!

A warm thanks also goes to all friends, neighbors, colleagues and students who volunteered as subjects in my experiments, sometimes lured by my home made apple cake, but mostly for nothing.

Last but not least: Thanks Anne, Axel and Björn for loving support and ability to maintain our home, your work and your studies in Sundsvall, as I was leading a somewhat wandering life.

# References

- Abercrombie, D. 1949. Teaching pronunciation. *English Language Teaching*, 3, 113-122
- Abrahamsson, N. & Hyltenstam, K. 2006. Inlärningsålder och uppfattad inföddhet i andraspråket – lyssnarexperiment med avancerade L2-talare av svenska. *Nordisk tidskrift för andrespråksforskning*, 1:1, 9-36.
- Ambrazaitis, G. 2007. Expressing 'confirmation' in Swedish: The interplay of word and utterance prosody. In *Proceeding of the XVI'th ICPHS*, Saarbrücken
- Anward, J. & Linell, P. 1976. Om lexikaliserade fraser i svenskan. *Nysvenska studier* 55- 56. 77-119.
- Aurén, J. A. 1869. *Bidrag till Svenska språkets ljudlära.*, Linköping
- Ballardini, K., Stjärnlöf, S. and Viberg, Å. 1997. *Nya mål 1*. Natur och Kultur, Stockholm
- Ballardini, K., Stjärnlöf, S. & Viberg, Å. 2001. *Nya mål 2*. Natur och Kultur, Stockholm
- Bannert, R. 1976. *Mittelbairische Phonologie auf akustischer und perzepto-rischer Grundlage*. Dissertation. Lund University. Travaux de l'Institut de Linguistique de Lund X.
- Bannert, R. 1979a. The effect of sentence accent on quantity. Invited paper to the *fifth symposium at the 9th international phonetics congress in Copenhagen*. Proceedings Vol. II, 253-259. Copenhagen
- Bannert, R. 1979b. Ordprosodi i invandrarundervisningen. I *Praktisk lingvistik nr 3*. Institutionen för lingvistik, Lunds universitet.
- Bannert, R. 1979c. Ordstruktur och prosodi. In: *Svenska i invandrarperspektiv : kontrastiv analys och språktypologi* : Hyltenstam, Kenneth (ed.). 2. uppl. – Lund : Liber Läromedel,.
- Bannert, R. 1980. Svårigheter med svenskt uttal: Inventering och prioritering. I *Praktisk lingvistik nr 5*. Institutionen för lingvistik, Lunds universitet.
- Bannert, R. 1984. Prosody and intelligibility of Swedish spoken with a foreign accent. *Nordic Prosody III. Acta Universitatis Umensis, Umeå Studies in the Humanities* 59, 7-18.
- Bannert, R. 1986. From prominent syllables to a skeleton of meaning: a model of prosodically guided speech recognition. In *Proceedings of the XIth ICPHS Tallinn*, pp 73-76.
- Bannert, R. 1990. *På väg mot svenskt uttal*. Studentlitteratur. Lund.
- Bannert, R. 1995. Intelligibility and acceptability in foreign accented Swedish: the effects of rhythmical and tonal features. *Reports from the Department of Phonetics, Umeå University, PHONUM* 3, 1995, 7-29.
- Behne, D. & Czigler, P. 1995. Distinctive vowel length and postvocalic consonant clusters in Swedish. *Phonum* 3, 55-63. Department of Phonetics, Umeå University.
- Behne, D., Czigler, P. & Sullivan, K. 1997. Swedish Quantity and Quality: A Traditional Issue Revisited. In *Phonum* 4, Dept of Linguistics, Umeå University.



- Behne, D., Czigler, P. & Sullivan, K. 1998. "Perceived Swedish vowel quantity: effects of postvocalic consonant duration" *International Conference on Spoken Language Processing*, December 1998, Sydney, NSW, Australia. 2291-2294.
- Bodén, P. & Grosse J. 2006. Youth Language in Multilingual Göteborg. *Proceedings of Fonetik 2006*. Department of Linguistics, Lund University.
- Boersma, P. & Weenink, D. 2001. *Praat – a system for doing phonetics by computer*. <http://www.fon.hum.uva.nl/praat/>
- Bohn, O-S. 1995. Cross-language speech perception in adults: First language transfer doesn't tell it all, in W. Strange (ed.) *Speech perception and linguistic experience: issues in cross-language research*, Baltimore, York Press: 273-304.
- Bongaerts, T., van Summeren, C., Planken, B. & Schils, E. 1997. Age and ultimate attainment in the pronunciation of a foreign language, *Studies in Second Language Acquisition*, 19, 447-465.
- Boyd, S. 2004. Utländska lärare i Sverige : attityder till brytning. In *Svenska som andraspråk: I forskning, undervisning och samhälle*. Hyltenstam, Kenneth and Lindberg, Inger (eds). Lund, Studentlitteratur, pp 419-436.
- Bruce, G. 1977. *Swedish word accents in sentence perspective*. Gleerup, Lund
- Bruce, G. 1987. On the phonology and phonetics of rhythm: evidence from Swedish. W. Dressler, H. Luschützky, O. Pfeiffer & J. Rennison (eds.) *Phonologica* 1984, 21-31. Cambridge University Press, London
- Bruce, G. 1998. *Allmän och svensk prosodi. Praktisk lingvistik nr 16*, Institutionen för lingvistik, Lunds universitet.
- Cebrian, Juli. 2007: Old sounds in new contrasts: l2 production of the English tense-lax vowel distinction. In *Proceedings of the XVI'th ICPHS, Saarbrücken*. 1637-1640
- Crystal, D. 2003. *A Dictionary of Linguistics and Phonetics*. Blackwell Publishing.
- Derwing, T. & Munro, M. (2005). Second language accent and pronunciation teaching: An empirical approach. *TESOL Quarterly*, 39, 379-397.
- Diehl, R. & Kluender, K. 1987. On the categorization of speech sounds. In *Categorical Perception: The Groundwork of Cognition*, edited by Steven Harnad, 226-253. Cambridge. Cambridge University Press.
- Elert, C-C. 1964. *Phonological Studies of Quantity in Swedish*. Uppsala: Almqvist & Wiksell.
- Elert, C-C. 1966. *Allmän och svensk fonetik*. Almqvist & Wiksell, Stockholm.
- Elert C-C 1970. *Ljud och ord i svenskan*. Almqvist & Wiksell, Stockholm
- Eliasson, S. 1985. Stress Alternations at a Vowel Length: New Evidens for an Underlying Nine-Vowel System in Swedish. *Nordic Journal of Linguistics* 8. 101-129.
- Elisasson, S. & La Pelle, N. 1973. Generativa regler för svenskans kvantitet. *Arkiv för nordisk filologi* 88 133-148
- Engstrand, O. 1988. Articulatory correlates of stress and speaking rate in Swedish VCV utterances. *Journal of the Acoustic Society of America* 83: 1863-1875
- Engstrand, O. 1995. Phonetic interpretation of the Word Accent Contrast in Swedish. *Phonetica* 52: 171-179
- Engstrand, O. 2004. *Fonetikens grunder*. Studentlitteratur, Lund: 116
- Engstrand, O. 2007. *Fonetik light*. Studentlitteratur, Lund: 39
- Eriksson, A. 1991. Aspects of Swedish speech rhythm. *Gothenburg monographs in linguistics 9*. Department of Linguistics, University of Göteborg.
- Eriksson, A. 2004. Swedia-projektet: dialektforskning i ett jämförande perspektiv. I: *Folkmålsstudier*, 43, 11–32. <http://swedia.ling.gu.se/>
- Escudero, P. & Boersma, P. 2004. Bridging the gap between L2 speech perception research and phonological theory. *Studies in L2 Acq.*, 26, 4: 551-585
- Fant, G. & Kruckenberg, A. 1989. Preliminaries to the study of Swedish prose reading and reading style. *STL-QPSR* 2, 1-83

- Fant, G. Kruckenberg, A. & Nord, L. 1991a. "Temporal organization and rhythm in Swedish", *Proceedings of the XIIIth ICPHS*, Aix-en-Provence, pp. 251-256.
- Fant, G. Kruckenberg, A. & Nord, L. 1991b. Durational correlates of stress in Swedish, French and English. *Journal of Phonetics* **19**, 351-365
- Fant, G. & Kruckenberg, A. 1994. Notes on stress and word accent in Swedish *STL-QPSR* 2-3/1994
- Field, J. 2005. Intelligibility and the listener: The role of lexical stress. *TESOL Quarterly*, 39, 399-423.
- Fischer-Jørgensen, E. 1964. "Sound Duration and Place of articulation" *Zeitschrift für Sprachwissenschaft und Kommunikationsforschung* 17:175-207
- Fischer-Jørgensen, E. and Jørgensen, H. P. 1969: Close and loose contact ("Anschluß") with special reference to North German. – In: *Annual Report of the Institute of Phonetics of the University of Copenhagen* (ARIPUC) 4, 43-80
- Flege, J. 1988. Factors affecting degree of perceived foreign accent in English sentences. *Journal of the Acoustical Society of America*, 84, 70-79.
- Flege, J. 1995. Second language speech learning: Theory, findings and problems. in W. Strange (ed.) *Speech perception and linguistic experience: Theoretical and methodological issues*. Timonium, MD: York Press.
- Flege, J. & Eefting, W. 1986. The production and perception of English stops by Spanish speakers of English. *Journal of Phonetics* **15**, 67-83.
- Flege, J. E., Munro, M. & MacKay, I. 1995. Factors affecting degree of perceived foreign accent in a second language. *Journal of the Acoustical Society of America* 97:3125-34.
- Flege, J., Bohn, O-S. & Jang, S. 1997. Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, 25, 437-470
- Flege, J. & Liu, S. 2001. The effect of experience on adults' acquisition of a second language. *Stud. Second Lang. Acquis.* 23, 527-552.
- Fraisse, P. 1982. Rhythm and Tempo. In Diana Deutsch (Ed.), *The Psychology of Music*, 149-180. New York: Academic Press.
- Fry, D. B. 1955. Duration and Intensity as Physical Correlates of Linguistic Stress. *The Journal of the Acoustic Society of America*, Volume 27, number 4: 765-768
- Garlén, C. 1988. *Svenskans fonologi*. Studentlitteratur, Lund.
- Gårding, E. 1974a. *Kontrastiv prosodi*. CWK Gleerups bokförlag, Lund.
- Gårding, E. 1974b. Den efterhängsna prosodin. I: *Teleman & Hultman Språket I bruk*. Liber, Lund.
- Gårding, E. 1979. Avvikande uttal – Analys och värdering. In *Praktisk Lingvistik 1*. Institutionen för lingvistik, Lunds universitet. 11-21.
- Gårding, E. & Lindblad, P. 1973. Constancy and Variation in Swedish Word Accent Patterns. *WP* 7, 36-110.
- Gårding, E., Bannert, R., Bredvad-Jensen, A-C., Bruce, G. & Naucclér, K. 1974. Talar skåningarna svenska? *Svenskans beskrivning 8. Förhandlingar vid sammankomst för att dryfta frågor rörande svenskans beskrivning* 8. sid. 107-117. Utgivare Platzack, Christer. Lunds universitet: Institutionen för nordiska språk.
- Gårding, E. & Bannert, R. 1979. Optimering av svenskt uttal. *Praktisk Lingvistik 1*. Institutionen för lingvistik, Lunds universitet.
- Hadding-Koch, K. & Abramson, A. 1964. Duration versus spectrum in Swedish vowels: Some perceptual experiments. In *Studia Linguistica* 18. 94-107.
- Hahn, L. D. 2004. Primary stress and intelligibility. Research to motivate the teaching of suprasegmentals. *TESOL Quarterly*, 38, 201-223.
- Heldner, M. & Strangert, E. 1998. On the amount and domain of focal lengthening in Swedish two-syllable words. *Proceedings of FONETIK 98*, Peter Branderund and Hartmut Traunmüller (eds.), Dept. of Linguistics, Stockholm University. 154-157.

- Heldner, M. & Strangert, E. 2001. Temporal effects of focus in Swedish *Journal of Phonetics*, 29 (3), 329-361.
- Higelin, S., Ekroth, G., Hjorth, A., Kristansen, S. & Tiegs, Å. 1973. *Samtal på svenska*. Sveriges Radios förlag.
- Holm, S. 2006. The Relative Contributions of Intonation & Duration to Degree of Foreign Accent in Norwegian as a second language. *Proceedings from Fonetik 2006*, Lund
- Huskisson, E. C. 1974. Measurement of pain *The Lancet* 9, 2, 1127-1131
- Håkansson, M. & Stenquist, A. 1989. *Om uttal*. Stockholm: Skriptor.
- Ingers, I. 1970. *Språket i Lund : dialekt och riksspråk i Lund och omnejd* / Stockholm: Sv.bokförl./Norstedts
- Jamieson, D. G. & Morosan, D. E. 1986. Training non-native speech contrasts in adults: Acquisition of the English /ð/-/θ/ contrast by Francophones. *Perception and Psychophysics* 40: 205-15
- Jamieson, D. G. and Morosan, D. E. 1989. Training new, non-native speech contrasts: A comparison of the prototype and perceptual fading techniques. *Canadian Journal of Psychology* 43: 88-96.0
- Johansson, K. 1981. Bör dubbelteckningsmetodikerna bygga på längd- eller klangfärgsskillnader? *Lund University. Lärarhögskolan i Malmö. Rapport 2*.
- Jonasson, J. & McAllister, R. 1972. Foreign accent and timing: An instrumental phonetic study. In *PILUS* (Papers from the Institute of Linguistics, University of Stockholm), 14, pp 11-40.
- Kjellin, O. 1976. A phonetic description of Tibetan; with a review of the literature. *Annual Bulletin* (Research Institute of Logopedics and Phoniatrics, University of Tokyo) (10), 127-143.
- Kjellin, O. 1977. Observations on consonant types and "tone" in Tibetan. *Journal of Phonetics*, 5, 317-338.
- Kjellin, O. 1978. 8:th edition 1995. *Svensk prosodi i praktiken*. Hallgren & Fallgren studieförlag, Uppsala
- Klatt, D. H. 1975. Vowel lengthening is syntactically determined in a connected discourse. *Journal of Phonetics* vol 3 pp 129-140
- Ladd, D. R., Mennen, I. & Schepman, A. 1999. The effect of vowel length and syllable structure on the location of pitch peaks in Dutch. *Architectures and Mechanisms in Language Processing*, Edinburgh, September 1999.
- Ladefoged P. & Maddieson I. 1996. *The Sounds of the World's Languages*. Oxford: Blackwell.
- Lehiste, I. & Peterson, G. E. 1959. "Vowel amplitude and phonemic stress in American English," *Journal of the Acoustical Society of America*, vol. 31, pp. 428-435.
- Lenneberg, E. H. 1967. *Biological foundation of language*. New York: Wiley.
- Liberman, M. and Prince, A. 1977. On stress and linguistic rhythm. *Linguistic Inquiry* 8, 249-336.
- Lindberg, I. 2006. *Språk och språkbruk bland ungdomar i flerspråkiga storstadsmiljöer 2000-2006*. Institute of Swedish as a Second Language, Göteborg University.
- Lindblom, B. Lyberg, B. & Holmgren, K. 1981. *Durational patterns of Swedish Phonology: Do they reflect short-term motor memory processes?*, Indiana Univ. Linguistic Club.
- MacKain, K. S., Best, C. T. & Strange, W. 1981. Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics* 2:369-90
- Maddieson, I. 1981 *UCLA Phonological Segment Inventory: Data and Index*. University of California. Working Papers in Phonetics/ 53 (Nov), pp. 1-242

- Malmberg, B. 1944. Die Quantität als phonetisch-phonologischer Begriff. In *Lunds universitets årsskrift N.F. Avd. I, 41:2*. Lunds universitet.
- Malmberg, B. 1956. Distinctive features of Swedish vowels: Some instrumental and structural data. In *For Roman Jakobson, 1956*. The Hague: Mouton, pp. 316-321. [Reprinted in Malmberg, B. 1971, *Phonétique générale et roman*. The Hague: Mouton, pp. 249-255]
- Malmberg, B. 1967. *Uttalsundervisning: teori och metodik*. Stockholm : Almqvist & Wiksell
- Marasek, K. 1996. Glottal correlates of the word stress and the tense/lax opposition in German *ICSLP 96. Proceedings., Fourth International Conference on Spoken Language Processing*. Volume 3, Issue , 3-6 Oct 1996 Page(s):1573 - 1576 vol.3
- Markham, D. 1997. *Phonetic Imitation, Accent, and the Learner (PhD thesis)*. (G. Bruce & Å. Viberg, eds., Travaux de l'Institute de Linguistique de Lund; vol. 33). Lund University Press.
- McAllister, R. 1997. "Perception and Production of a Second Language and the Concept of Foreign Accent" i Leather J & James A (eds.): *New Sounds 97*. University of Klagenfurt: 206-214.
- McAllister, R., Flege, J.E. & Piske, T. 1999. The Acquisition of Swedish Long vs. Short vowel Contrast by Native Speakers of English, Spanish and Estonian. In J. Ohala, Y. Hasegawa, M. Ohala, D. Granville & A. Bailey (Eds.) *Proceedings of the XIVth International Congress of Phonetics Sciences* (Berkeley, CA: Department of Linguistics, University of California at Berkeley), pp. 751-754.
- McAllister, R. 2001. Experience as a factor in L2 phonological acquisition. In the *Proceedings of Fonetik 2001, the XIVth Swedish Phonetics Conference*, Örenäs, May 30 - June 1, 2001. Working Papers, Department of Linguistics, Lund University 49: 116-119.
- McAllister, R., Flege, J. & Piske, T. 2002. The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. *Journal of Phonetics* 30, 229-258.
- McAllister, R., Inoue, M. & Dahl, S. 2006. L1 Residue in L2 Use: A Preliminary Study of Quantity and Tense-lax. *Proceedings from Fonetik 2006 Lund*, pp. 93-96.
- Missaglia, F. 1999. Contrastive prosody in SLA – an empirical study with adult Italian learners of German, *Proceedings of the 14<sup>th</sup> International Congress of Phonetic Sciences*, Vol. 1, pp 551-554.
- Morosan, D. E. & Jamieson, D. G. 1989. Evaluation of a technique for training new speech contrasts: Generalization across voices, but not word position or task. *Journal of Speech and Hearing Research* 32:501-11.
- Moyer, A. 1999. Ultimate attainment in L2 phonology. The critical factors of age, motivation and instruction. *Studies in second language Acquisition*, 21, 81-108.
- Munro, M. J. 1995. Nonsegmental factors in foreign accent: ratings of filtered speech. *Studies in Second Language Acquisition*, 17, 17-33.
- Munro, M. J. & Derwing T. M. 1995. Foreign accent, comprehensibility and intelligibility in the speech of second language learners. *Language Learning*, 49 (Supp. 1), 285–310.
- Munro, M. & Derwing, T. (1999). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, 49, Supplement 1, 285-310.
- Munro, M. J. & Derwing, T. M. 2005. Second Language Accent and Pronunciation Teaching: A Research-Based Approach. In *Tesol Quarterly* volume 39, nr 3: 379-397

- Möbius, B. 2003. Gestalt Psychology Meets Phonetics – An Early Experimental Study of Intrinsic F0 and Intensity. *Proc. XVth ICPHS*, Barcelona.
- Piske, T., MacKay, I. & Flege, J. E. 2001. Factors affecting degree of foreign accent in an L2: a review. *Journal of Phonetics* 29, 191-215
- Purcell, E. & Suter, R. 1980. Predictors of pronunciation accuracy: A reexamination, *Language Learning*, 30, 271-287.
- Reuter, M. 1971. Vokalerna i finlandssvenskan – En instrumentell analys och ett försök till systematisering enligt särdrag. *Studier i nordisk filologi* 58, sidan 240-249. Helsingfors.
- Reuter, M. 1982. *Kvantitet i Helsingforssvenskan*. Unpublished dissertation for Licentiate of Philosophy degree, University of Helsinki.
- Riad, T. 1992. *Structures in Germanic Prosody*. Doctoral dissertation. Stockholm: Stockholm University.
- Rochet, B. 1995. Perception and production of second-languages speech sounds by adults. In: W. Strange (ed.) *Speech perception and linguistic experience: issues in cross-language research*, Baltimore, York Press: 379-410.
- Schaeffler, F. 2005. *Phonological quantity in Swedish dialects: A data-driven categorization*. Doctoral thesis, Department of phonetics, Umeå University.
- Slagbrand, Y. & Thorén, B. 1982. *Övningar i svensk basprosodi*. Local teaching media at the University of Jyväskylä, Finland.
- Slagbrand, Y. & Thorén, B. 1997. *Övningar i svensk basprosodi. Lärarhandledning och övningar i svenskt uttal inom sfi, svenska som andraspråk och svenska som främmande språk*. Semikolon; Boden.
- Sluijter, A. M. C. & van Heuven, V. J. 1996. Acoustic correlates of linguistic stress and accent in Dutch and American English. *Proceedings of the fourth International Conference on Spoken Language Processing (ICSLP)*. Philadelphia
- Strange, W. 1995. "Phonetics of Second Language Acquisition: Past, Present and Future" i *Proceedings ICPHS 95*. Vol 5. 76-83.
- Strangert, E. 1985. *Swedish Speech Rhythm in a cross language perspective*. Doctoral thesis. Department of Phonetics, Umeå University.
- Strangert, E. & Wretling, P. 2003. "Complementary quantity in Swedish dialects" In *Proceedings of Fonetik 2003*, Umeå/Löfvånger. 101-104
- Sundberg, U. 1999. Quantity in Infant-Directed Speech. In *Proceedings of the XIVth International Congress of Phonetic Sciences*. San Fransisco. 2189-2191
- Tajima, K., Port, R., & Dalby, J. 1997. Effects of temporal correction on intelligibility of foreign-accented English. *Journal of Phonetics*, 25, 1-24.
- Thompson, I. 1991. Foreign accent revisited: The English pronunciation of Russian immigrants, *Language Learning*, 41, 177-204.
- Thorén, B. 1982. *Svenskans rytm*. Unpublished BA paper. Department of Linguistics, Stockholm university.
- Thorén, B. 1988. *Betoningshandboken – liten hjälprea för oss som undervisar i svenska som andraspråk*. BT Bättre svenska, Stockholm. 2:nd edition 1994, Sundsvall.
- Thorén, B. 2001. *Vem vinner på längden? Två experiment med manipulerad duration i betonad stavelse*. Opublicerad D-uppsats i fonetik. Institutionen för filosofi och lingvistik, Umeå universitet.
- Thorén, B. 2004. The postvocalic consonant as a complementary cue to the perception of quantity in Swedish. *Proceedings from FONETIK 2004*. Institutionen för lingvistik, Stockholms universitet.
- Thorén, B. 2006. Phonological demands vs. system constraints in an L2 setting. In *Proceedings from FONETIK 2006*. Institutionen för lingvistik, Lunds universitet.

- Thorén, B. 2008. (in press). Funderingar och forskning kring en basprosodisk undervisningsstrategi för uttalsundervisning i svenska som andraspråk. *Proceedings from ASLA 2006*. Mälardalens högskola, Eskilstuna.
- Traunmüller H. & Bigestans A. 1988. Perception of the quantity distinction in Swedish /VC/-sequences. *Papers from the Second Swedish Phonetics Conference held at Lund, May 5-6 1988* 124-127
- Wang, X. & Munro, M. J. 1999. The perception of English tense-lax vowel pairs by native Mandarin speakers: The effect of training on attention to temporal and spectral cues. *Proc. 14th ICPhS*, 125-129. Berkeley, CA: University of California
- Wewers, M. & Lowe, N. 1990. A critical review of visual analogue scales in the measurement of clinical sciences. *Research in Nursing and Health*, 13, 227-236.
- Yamada, R. A. & Tokhura, Y. 1992. Perception of American English /r/ and /l/ by native speakers of Japanese. In Y. Tokhura, E. et al. (Eds.) *Speech perception, production and linguistic structure*. Tokyo, Japan: OHM Publishing Co. Ltd. 155-174.