Chapter 2

Background

In this chapter the linguistic and theoretical background for the thesis is presented. In § 2.1 the status of dialects and regional varieties of Standard Swedish in Sweden and the Swedish-language parts of Finland is described. In § 2.2 different classifications of the varieties of Swedish are presented; § 2.2.1 shows how the rural dialects have been classified, while § 2.2.2 shows the main regional varieties of Standard Swedish. Since the data for the present study comes from the SweDia database, classifications made based on some specific linguistic features within the SweDia project are described in § 2.2.3.

In § 2.3.1 the Swedish vowel system is described. Variation in the pronunciation of the Standard Swedish vowel phonemes is described in § 2.3.2, while § 2.3.3 shortly covers the main sources of variation in the very diverse vowel systems in the traditional rural dialects.

As a background for the choice of acoustic methods for this thesis, different methods for measuring vowel quality acoustically are discussed in § 2.4. Measurement of formants is described in § 2.4.1, while different whole-spectrum approaches are discussed in § 2.4.2. The problem of speaker variability in acoustic measurements of vowels is explained (§ 2.4.3) and different solutions for normalizing for the speaker-dependent variation are discussed (§ 2.4.4).

In § 2.5 different methods used in dialect geography and dialectometry are presented as a basis for the choice of methods for analyzing dialectal variation in this thesis.

2.1 Swedish dialects and standard language

Swedish is a North Germanic language spoken as a first language by around nine million speakers. Out of these approximately 8.5–9 million live in Sweden (Börjars, 2006) and nearly 300,000 in Finland (Statistics Finland). Swedish is the main language in Sweden and one of two official languages in Finland. The Swedish population in Finland lives mainly along the coasts and comprises 5.4% of the Finnish population (Statistics Finland).
Chapter 2. Background

The Swedish written language was standardized during the 18th century (Teleman, 2005). The spoken standard language developed in the capital and was influenced by the court and the speech of higher social classes in Stockholm. By the 20th century the spoken and written standard language was well established (Thelander, 2009). The regional variation in Standard Swedish today is generally on the prosodic level, primarily intonation, but there is also variation on the sub-phonemic level, especially in vowel pronunciation. Local dialects show variation from Standard Swedish on all linguistic levels and are not always intelligible for speakers of Standard Swedish.

2.1.1 Standard Swedish

The Swedish standard language is well-codified with a number of relatively recent dictionaries and handbooks\(^1\) (Thelander, 2009). The Standard Swedish written language is uniform and there is a norm which is accepted in the whole language area. For the spoken language, however, a neutral standard variety, which would not be geographically identifiable, hardly exists (Garlén, 2003, 7–8). Rather, there are a number of regional varieties of Standard Swedish, which differ from each other when it comes to prosodic features and the pronunciation of certain phonemes. In the pronunciation dictionary *Svenska språknämndes uttalsordbok* a rather broad definition of Standard Swedish pronunciation is given: Standard Swedish pronunciation is defined as a pronunciation which can be generally accepted and used in the whole language area (“en uttalsform som kan accepteras och brukas allmänt över hela det svenska språkområdet”) (Garlén, 2003, 7). The dictionary is not strictly normative, but the aim is to give recommendations which can be applied to most of the varieties of Standard Swedish. For many words several pronunciations are given in the dictionary. The aim is not to create a uniform spoken language, but to recommend forms which lead to good intelligibility in all parts of the language area (Garlén, 2003, 8).

In comparison to the relatively lax attitudes towards variation in Standard Swedish nowadays, exemplified by the definition of Standard Swedish in the pronunciation dictionary, the attitudes were much more rigid up until the 1970s. For example, there was a demand for news anchors in television and radio to speak neutral Standard Swedish, that is, a standard variety which did not signal geographic provenance (Thelander, 2009). Because Standard Swedish had developed in the capital, the neutral Standard Swedish was affiliated with the Central Swedish speech tradition around Lake Mälaren (close to Stockholm and Uppsala). This Central Swedish pronunciation is the one that is still used in most schematic descriptions of Swedish.

Since the 1970s the attitudes have become more tolerant and nowadays it is not uncommon to hear news anchors with marked local features in their pronunciation or a foreign accent (Thelander, 2009). According to Svensson (2005), one reason for the increased linguistic variation in the public domain is the increase in the length of compulsory education. Until the 1960s, most of the Swedes followed only six years of elementary school. Today, almost everyone follows at least eleven or twelve years of schooling, including non-compulsory upper secondary school. With the higher education level, a growing number of people participate in public debate and thereby the linguistic variation in public language has increased and the standard norm has weakened. Thelander (2009) argues that discussions about correct norms have become superfluous, because a majority of the Swedes feel secure in their use of the (regional) standard language.

The Swedish spoken in Finland follows the same standard norms as Swedish in Sweden (that is, Central Swedish). However, the language-contact situation with Finnish continuously influences the Swedish spoken and written in Finland. Moreover, the fact that a language is used in two different countries with different societies and social systems naturally leads to some differences. According to Thelander (2009), not only the spoken language, but also the written language of Swedish speaking Finns can be recognized relatively easily by Swedes. In addition to Finnish, the regional standard variety of Swedish spoken in Finland has also been influenced by the Finland-Swedish dialects and some other languages (for example, Russian, German). Finland-Swedish also employs some features that are considered archaic in Sweden. Swedish-speaking Finns who live in areas dominated by Finnish are usually bilingual, and under the right circumstances individuals develop a balanced bilingualism with high proficiency in both Swedish and Finnish (Tandefelt, 1996). However, the extent to which Swedish-speaking Finns have the opportunity to use Swedish in their daily lives has been shown to determine the development of high proficiency in Swedish and idiomatic use of the Swedish language (Leinonen & Tandefelt, 2000, 2007).

2.1.2 From dialect diversity to leveled dialects

Most European languages have recently gone through processes of dialect leveling. Auer (2005) has showed that, in spite of superficial heterogeneity in the dialect-standard constellations found in Europe, the chronological development from local base dialects to a spoken standard variety with only little variation can be described systematically with a few types. Industrialization has played on important role in this leveling of the base dialects.

Swedish has shown extensive geographical variation from medieval times until the first half of the 20th century (Hallberg, 2005). More or less every parish had a dialect of its own, distinct from the neighboring dialects. These rural dialects were characterized by dialectal features at all linguistic levels: segmental, phonology, prosody, morphology, lexicon, semantics, syntax. The 20th century changed this linguistic
situation dramatically. The main causes were industrialization, urbanization and migration, all naturally connected to each other.

Industrialization started in Sweden around 1870. Around 1900 more than half of the working population of Sweden was still employed in agriculture, but at the beginning of the 21st century agriculture employed less than three percent (Thelander, 2005, 1905). The industrialization resulted in rapidly growing industrial communities and lead to a “flight from the countryside”. In 1850 90% of the Swedish population still lived in the countryside, in 1980 only 17% (Hallberg, 2005, 1691). Industry grew until the 1960s. After that, it, too, lost importance and today two thirds of the population are employed in the service sector (Thelander, 2005, 1905). In this societal shift, the rural life style, which had been dominant in earlier days, was almost completely replaced by an urban life style (Nordberg, 2005, 1759).

The linguistic result of the societal shift was large-scale homogenization. In the cities and industrial communities, the dialects of immigrants mixed and became simplified. Examples of simplifications of the dialects are a replacement of the grammatical three gender system by a two gender system and the loss of vowel phonemes when diphthongs merge with long vowels. In communities with local dialects very divergent from the standard language, the immigrants did not always learn the local variety but spoke Standard Swedish instead, and, hence, diffused features from the standard language into the local dialect (Nordberg, 2005, 1769). When keeping the contact with their original social networks, migrants also, to some extent, contributed to the diffusion of standard variants in the locality they had moved from (Nordberg, 2005, 1769).

Apart from migrations, the school system also added to the dialect leveling. Especially in areas with divergent rural dialects “the local variety was counteracted at school up until the 1970s at least, even if it was seldom forbidden” (Nordberg, 2005, 1767).

One of the most important changes in the linguistic situation in Sweden during the past century is that while earlier many people grew up in a code-switching situation between dialect and standard language, today code-mixing best describes the language situation for the majority of the Swedish-speaking (Andersson, 2007, 55). Today, the linguistic distance between local varieties and the standard language is generally so small that the two varieties cannot be seen as separate linguistic systems, but speakers make use of a gliding scale where the share of dialectal features and standard variants vary according to speech situation, speech partner and the degree of formality.

Swedish spoken language can be categorized as belonging to one of the four following levels (Thelander, 1994; Hallberg, 2005):

- rural local dialect
- regional, leveled dialect
- regional standard language
- neutral standard language
As mentioned in the previous section a neutral Standard Swedish, which does not signal geographic provenance, hardly exists. The varieties regional, leveled dialect and regional standard language were results of the language homogenization of the past century. In this process, the most local, divergent dialectal features were lost, but features representative for a larger region persisted. This regionalization of local dialects was shown clearly in a study by Thelander (1979) of the dialect of Burträsk in Västerbotten. The larger the geographic spread of a dialectal feature in the surroundings of Burträsk, the more prone the Burträsk subjects were to use the dialect variant instead of the standard variant.

Auer (2005) calls the stage with intermediate varieties between standard language and local dialects a diaglossic repertoire (in contrast to the diglossic repertoire where speakers code-switch between their local dialect and a spoken standard language). Both a diaglossic and a diglossic situation can lead to the loss of local dialects.

Today, at the beginning of the 21st century, the rural local dialects, are disappearing in Sweden, and most speakers are found somewhere on the scale between regional dialect and regional standard language. Only in some peripheral areas (especially Upper Dalarna, Norrbotten and Gotland) local dialects are still spoken. In these areas, the local dialect and Standard Swedish are perceived as two separate linguistic systems and the speakers are bidialectal. In many of these places, however, speakers of the local dialect are found mainly among older people.

It may seem contradictory that rural dialects are disappearing in times when the attitudes towards linguistic variation are relatively liberal (compare § 2.1.1). However, according to Auer this is not uncommon: “In the final stage before loss, the attitudes towards the now almost extinct base dialect are usually positive again, and folkloristic attempts at rescuing the dialect may set in – usually without success” (Auer, 2005, 29). In a study in Överkalix (Norrbotten) in 1988, Källskog (1990) found overall positive attitudes towards the local dialect among junior high school students. 78% of the dialect-speaking students and 66% of the students who did not speak the local dialect but only Standard Swedish had positive attitudes towards the local dialect. The local dialect was affiliated with belonging to the home district, and being able to talk to one’s grandparents was mentioned as something positive about the dialect. However, only 36% of all junior high school students (grades 7–9) in Överkalix were actually speakers of the local dialect. Of their parents, 70% spoke the dialect, which shows a remarkable decline. It turned out that many of the parents had experienced negative attitudes towards the dialect in their youth or had had to abandon the dialect when entering higher education. Many parents had chosen to speak Standard Swedish to their children in order to avoid problems at school or to make it easier for the children to get an education and job outside Överkalix later on. The attitudes of the previous generation, hence, were decisive for the declined use of dialect in the younger generation.

In a study of attitudes towards a number of Swedish dialects by Bolhek Radovani (2000), 42% of the subjects answered that dialect can be used in all circumstances, which seems very liberal. However, the study also showed that the subjects interpreted the word “dialect” differently than linguists would. The varieties that the
subjects considered dialects would be categorized as leveled dialect or regional standard language by Swedish linguists. This fits in with Auer’s typology according to which the repertoire is restructured when the rural dialects are lost, and the regiolectal forms are now considered the “most basilectal way of speaking” (Auer, 2005, 27).

Even in the far-reaching processes of linguistic homogenization, differences between rural and urban communities do still exist. In the cities and towns, there is more social and linguistic stratification than in rural areas. A large number of studies of Swedish urban and rural communities have shown that higher social groups generally use more standard variants, while local variants are preferred by lower social groups. In rural environments, less socio-economically defined linguistic variation is found, because the population is more homogeneous. In rural settings, the networks are generally also smaller and more close-knit, which influences the linguistic behavior (Nordberg, 2005, 1761).

The Swedish dialects in Finland have had a stronger position throughout the 20th century than the dialects in Sweden (Reuter, 2005, 1655). One of the reasons for this is that industrialization reached Finland somewhat later than Sweden, starting in the 1880s, and was slower in the initial phase. Until the 1960s the majority of the Finnish population still lived in a rural environment (Tandefelt, 1994). Another reason is that elementary school was not introduced in Finland until the 1920s (Reuter, 2005, 1655). In the Swedish language area in Finland, especially in the province Österbotten, bidialectalism is still common, and the majority of the speakers have a local dialect as their first language. This holds not only for the countryside but also for smaller towns (Ivars, 1996).

However, in spite of positive attitudes towards local dialects, regionalization has also affected the Swedish dialects in Finland to some extent. The regionalization tendencies are stronger in the southern parts of the Finland-Swedish area (above all close to Helsinki) than in Österbotten (Ivars, 2003; Sandström, 1996). In the region close to the capital in Finland, rural dialects have disappeared not only because of change towards Standard Swedish, but also because of language shift to the majority language Finnish (Tandefelt, 1988, 1994, 1996).

### 2.2 Swedish dialect geography

Since the 1930s numerous dialect geographic works, including maps, have been published describing the dialects in the Swedish language area (for an overview see Edlund, forthcoming). However, no comprehensive dialect atlas covering the whole language area has yet been compiled. The existing atlas works include only smaller parts of the language area, while dialect geographic works including the whole Swedish (or Nordic) dialect area are generally monographs dealing with some specific features or words. In the Swedish language area, there is a stronger tradition

---

for compiling dialect dictionaries than dialect atlases. A number of dictionaries and word lists covering smaller or larger dialect areas of the Swedish language area exist. Of the dialect dictionary covering the dialects of Sweden—Ordbok över Sveriges Dialekter (Reinhammar & Nyström, 1991–2000)—unfortunately only three booklets have been published covering words in the range A–back. At the moment, no more booklets of the work are being published, but the archive put together for compiling the dictionary is open for researchers. The dictionary covering the Swedish dialects in Finland—Ordbok över Finlands svenska folkmål (Ahlbäck & Slotte, 1976–2007)—has reached the word och and is being compiled at the Research Institute for the Languages of Finland.

Lexical geography in the Wörter-und-Sachen tradition has been especially strong in the Swedish language area resulting in several monographs concerning a particular word or semantic field. These lexical studies have provided insight into phonological history and change, etymology and semantic development (Edlund, forthcoming). Numerous dialect geographic studies concerning phonetics and phonology also exist. These have often dealt with sound changes from a historical point of view. Mappings of morphology, syntax and prosody are less frequent, but a smaller number of studies dealing with these linguistic levels exist.

Examples of more extensive dialect geographic works concerning specific regions of the Swedish language area are the work in five volumes by Götlind & Landtmanson (1940–50) dealing with the dialects of Västergötland, Södschwedischer Sprachatlas by Benson (1965–70) and a dialect atlas of the northern part of Norrland by Hansson (1995).

Standard works describing the Swedish dialects are Våra folkmål by Wessén (1969), which was first published in 1935, and Svenska dialekter by Pamp (1978). Pamp describes the dialects in each of the historical provinces of Sweden without suggesting any linguistic classification of the dialects, beyond the administrative province borders. Figure 2.1 displays the Swedish provinces. The provinces of Sweden are grouped as belonging to one of the three larger regions Götaland, Svealand and Norrland. Reference to the province names are used when discussing the results in the following chapters of this thesis.

Pamp (1978) does not deal with the Swedish dialects in Finland. Descriptions of the Finland-Swedish dialects are found in Svenskan i Finland by Ahlbäck (1956) and Från Pyttis till Nedervetiit by Harling-Kranck (1998). Within the Finland-Swedish area a division according to provinces is usually applied. The provinces with Swedish population in Finland are displayed in Figure 2.1.

Wessén (1969) suggested a linguistic classification of the Swedish dialects into six groups. This classification is described more closely in § 2.2.1. Elert (1994) proposed a division of the regional varieties of Standard Swedish that largely resembles the classification of the traditional rural dialects by Wessén. Elert’s classification of the regional varieties of Standard Swedish is described in § 2.2.2.

Recently, data collected in the SweDia project (see § 4.1) have made it possible to conduct quantitative analyses of modern spoken Swedish. The data was gathered around year 2000, and within the project, Swedish dialects have been classified ac-
Figure 2.1. The historical provinces of Sweden and Swedish-speaking parts of Finland. Sweden is divided into the three larger regions Götaland, Svealand and Norrland.
2.2. Swedish dialect geography

According to some specific linguistic features. Since the data for this thesis comprises vowel data from the SweDia database, an overview of classifications made based on data of other linguistic levels from the same database is given in § 2.2.3. Comparing the results of different studies where SweDia data have been used makes sense because all the studies involve the same participants recorded at the same point in time. The studies of variation at different linguistic levels are therefore directly comparable.

In Swedish dialectology, computational methods have not been very commonly used (Edlund, forthcoming). In some of the studies described in § 2.2.3, methods borrowed from the dialectometric research tradition (see § 2.5) have been applied (particularly cluster analysis), and Leinonen (2007) used cluster analysis and multidimensional scaling for analyzing vowel pronunciation in Finland-Swedish dialects. However, Swedish dialect geographic works where quantitative methods are used have generally focused on some specific part of the language system, rather than aiming at an aggregate analysis, which is what has been the main focus of dialectometry. Aggregate dialectometric analyses of the whole Swedish language area do not exist so far. In the dialect atlas of the northern part of Norrland (Hansson, 1995) some summarizing dialectometric maps are included.

2.2.1 Classification of Swedish rural dialects

According to Wessén (1969, 12–13), the rural Swedish dialects have formed a continuum without any sharp dialect borders. Neither has this continuum been disrupted by national borders in Scandinavia; Danish and Norwegian dialects belong to the same continuum. Even though Wessén recognized that no abrupt borders existed between dialect areas, he motivated a classification of the dialects with practical reasons: a sketch of a dialect division will help to give an overview of the varying linguistic phenomena. Wessén (1969) described the Swedish dialects as belonging to six main dialect areas:

- South Swedish dialects (sydsvenska mål)
- Götaland dialects (götamål)
- Svealand dialects (sveamål)
- Norrland dialects (norrländska mål)
- Gotland dialects (gotländska mål)
- Finland-Swedish dialects (östsvenska mål)

Figure 2.2 (left) shows the approximate areas. The Svealand dialects are divided into three sub-groups: East Central Swedish (Sw. uppsvenska), Middle Central Swedish (Sw. mellansvenska) and the dialects of Dalarna (Sw. dalmål). The classification is based mainly on phonetic, phonological and morphological features, viewed from a historical perspective. The division has been commonly used by Swedish dialectologists.
Figure 2.2. The map to the left shows the classification of Swedish rural dialect according to Wessén (1969), while the map to the right shows the division of modern spoken Swedish proposed by Elert (1994). The divisions are rather similar.

The South Swedish area includes dialects in the provinces Skåne, Blekinge and southern parts of Halland and Småland.

The province Västergötland is the center of the Götaland dialects. Other provinces that Wessén includes in the Götaland area are Dalsland, northern Småland, northern Halland and the south-west of Östergötland. Värmland is also included in the Götaland area, even though it has a special status when it comes to many features. Bohuslän is a transitional area between South Swedish, Götaland and Norwegian dialects. Some Götaland features have spread via Värmland and western parts of Västmanland to the north.

The center of the Svealand dialects is in Uppland. Uppland together with Gästrikland, south Hälsingland, south-east Dalarna, eastern parts of Västmanland and northern and eastern parts of Södermanland form the East Central Swedish area. Närke and the rest of Södermanland form a transitional area between Svealand and Götaland dialects, and the same is true for Östergötland, north-east Småland and Öland. These dialects are called Middle Central Swedish. The dialects in Dalarna
are very conservative and divergent and comprise a separate group within the Svealand dialects.

The Norrland dialects are spoken along the Swedish coast from north Hälsingland to Norrbotten in the north. Additionally, the dialects in Jämtland can be included in the Norrland area, even if they share a number of features with Norwegian dialects. The dialects in Härjedalen and north-west Dalarna are of a Norwegian type, according to Wessén.

The Swedish dialects in Finland have comprised an East Swedish dialect area together with the Swedish dialects in Estonia (spoken along the western coast of Estonia). However, most of the Swedish population of Estonia fled to Sweden during the Second World War and Estonian-Swedish dialects are almost extinct today. The Finland-Swedish dialects share many features with the East Central Swedish ones.

The dialects on the island Gotland have preserved many conservative features and are very different from mainland-Swedish ones.

### 2.2.2 Regional varieties of Standard Swedish

The division of the regional varieties of Standard Swedish proposed by Elert (1994) is displayed in Figure 2.2 (right). A comparison of the division of rural Swedish dialects by Wessén (1969) in the same figure shows many similarities. Elert (1994) proposed three main varieties of Standard Swedish: South Swedish, Central Swedish and Finland-Swedish. Central Swedish was further subdivided resulting in a division into seven groups:

- South Swedish (sydsvenskt talspråk)
- East Central Swedish (öst mellansvenskt talspråk)
- West Central Swedish (götiskt-västmellansvenskt talspråk)
- the spoken language of Bergslagen (bergslagstalspråk)
- the spoken standard language of Norrland (norrlandsstandardsvenskt talspråk)
- the spoken language of Gotland (gotlänstal språk)
- Finland-Swedish (finlandssvenskt talspråk)

Elert (1994) based his division mainly on sentence intonation and differences in vowel pronunciation. In addition, some varieties are characterized by salient features like the use of dorsal /r/ in South Swedish or the lack of the word accent distinction in Finland-Swedish. The regional variation in vowel pronunciation in Standard Swedish is discussed in more depth in § 2.3.2.

### 2.2.3 Typologies based on specific features

Based on data from the SweDia database (see § 4.1), Bruce (2004) classified Swedish dialects according to intonational variation. The intonational parameters of the model were focal accentuation, phrasing, word accentuation and compounding. Seven
distinct dialect regions were identified, largely corresponding to the ones found by Elert (1994) (§ 2.2.2). Bruce (2004) called the seven intonational types South, West, Central, East, Far East S, Far East N and North. Far East stands for Finland-Swedish dialects, which were divided into two subtypes: Southern (the south coast) and Northern (the west coast).

Schaeffer (2005) also used data from the SweDia database. He used cluster analysis for classifying the Swedish dialects based on phonetic variation in quantity. Schaeffer found a division into three main types: Southern Swedish (up to Uppland and Middle Dalarna), Northern Swedish (Norrland and Åland) and Finland-Swedish (mainland Finland-Swedish dialects). The three areas are separated mainly by consonant length. The Finland-Swedish dialects are characterized by a shorter consonant than the two other areas in V:C sequences, while the Southern Swedish area shows a markedly short consonant in VC: sequences. In the Northern area more preaspiration was found than in the two other areas. The phonetic differences between the three types could be connected to the phonological systems of the dialects, that is, the presence or absence of VC and V:C: syllables in stressed positions (see § 2.3.3).

Lundberg (2005) used clustering methods for analyzing differences in the pronunciation of the vowel in the word lat in Swedish dialects analyzed with Mel-frequency cepstral coefficients (MFCCs, see § 2.4.2). The data of the study comprised older male speakers from the SweDia database. The geographic variation in the Swedish vowel a elicited with the word lat was studied and three clusters representing different variants of the vowel were found. The study showed that clustering should not be applied without evaluation. Principal component analysis was used to establish which MFCCs were important for identifying the clusters.

2.3 Swedish vowels

The Standard Swedish vowel system is described below in § 2.3.1. As mentioned above there is no neutral, geographically and socially non-identifiable, Standard Swedish, but a number of regional varieties of Standard Swedish exist. In § 2.3.2 the regional differences in vowel pronunciation are described.

The vowel systems of Swedish rural dialects differ from Standard Swedish both phonetically and phonologically. In § 2.3.3 an overview of some general features concerning a number of Swedish dialects is given.

2.3.1 The Standard Swedish vowel system

Standard Swedish has eighteen vowel phonemes, nine long and nine short ones corresponding to each other pairwise. Table 2.1 displays these vowels and their Swedish orthographic equivalents. The correspondence between the long and short vowels is not only an orthographical one, but the correspondence exists in the linguistic competence of native speakers and is founded on phonetic similarity as well as morphophonological alternations (Linell, 1973, 8).
### Table 2.1. The Standard Swedish vowel phonemes (Hedelin, 1997).

<table>
<thead>
<tr>
<th>Swedish letter</th>
<th>Long vowel</th>
<th>Short vowel</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>/aː/</td>
<td>/a/</td>
</tr>
<tr>
<td>e</td>
<td>/ɛː/</td>
<td>/ɛ/</td>
</tr>
<tr>
<td>i</td>
<td>/iː/</td>
<td>/i/</td>
</tr>
<tr>
<td>o</td>
<td>/uː/</td>
<td>/o/</td>
</tr>
<tr>
<td>u</td>
<td>/uː/</td>
<td>/ø/</td>
</tr>
<tr>
<td>y</td>
<td>/yː/</td>
<td>/γ/</td>
</tr>
<tr>
<td>å</td>
<td>/oː/</td>
<td>/o/</td>
</tr>
<tr>
<td>ä</td>
<td>/ɛː/</td>
<td>/ɛ/</td>
</tr>
<tr>
<td>ö</td>
<td>/øː/</td>
<td>/ø/</td>
</tr>
</tbody>
</table>

The phonemes are displayed with IPA symbols used for Standard Swedish in the pronunciation dictionary *Norstedts svenska uttalslexikon* (Hedelin, 1997). The pronunciation dictionary *Svenska språknämndens uttalsordbok* (Garlén, 2003) appeared a few years later, but does not use IPA symbols for all vowels, which is the reason why symbols in *Norstedts svenska uttalslexikon* were chosen to denote the Standard Swedish vowel phonemes and Standard Swedish pronunciation throughout this thesis.

A few differences exist between the two mentioned Swedish pronunciation dictionaries. The long a, transcribed [aː] by Hedelin (1997), is transcribed [uː] in the more recent dictionary. Garlén (2003, 31) describes the Swedish long a vowel as a slightly rounded open back vowel. Hence, the actual pronunciation is something between unrounded [aː] and rounded [uː].

The other vowel for which the two dictionaries have used different IPA symbols is the short ö. Hedelin (1997) uses [o], while Garlén (2003) uses [œ]. According to Garlén (2003) the pronunciation is, thus, more open than according to Hedelin (1997). However, according to both authors the pronunciation of short ö is more open than the pronunciation of long ö, which is probably more important than the exact degree of openness of the vowel.

According to Elert (1997) in the Introduction to *Norstedts svenska uttalslexikon*, the Swedish long o is somewhat more open than the cardinal vowel [u]. Therefore the phonetic symbol [œ] is used in *Norstedts svenska uttalslexikon*. The symbol [o] denotes a semi-high back rounded vowel, but was marked as obsolete by IPA in 1989 and is not used in the newest version of the International Phonetic Alphabet. Throughout this thesis [uː] is therefore used for the Standard Swedish long o.

The pronunciation of long u in Standard Swedish is actually more fronted than the symbol [u] suggests. According to Elert (1997) a more precise phonetic symbol would be [ʉː].

In addition to the eighteen phonemes, more open allophones of ä and ö are used when these vowels are followed by /r/ (which is equal to [r] or a retroflex consonant resulting from the consonant combinations rd, rt, rn, rs and rt):
In an acoustic analysis of Swedish long vowels, Eklund & Traunmüller (1997) showed that only /æ:/ and [æ:] are plain monophthongs. Especially the mid vowels, /ø:/, /øː/ and /e:/, showed substantial diphthongization. These vowels are pronounced as opening diphthongs; F1 increases in the course of the vowels, and F2 decreases in the front vowels, /ø:/ and /æ$/, and increases in the back vowel /øː/. The close vowels (/iː/, /yː/, /uː/ and /uː:/) showed smaller formant movements than the mid vowels. The close vowels first become even more close and only at the very end more centralized. The subjects (six male and six female in the age range 20–58) were from the greater Stockholm area.

Describing the Swedish vowels with distinctive features has turned out to be complicated. A simple matrix, which in addition to the length distinction includes three degrees of openness, a front–back distinction, and a roundness distinction for front vowels, can be set up (Table 2.2). The problem is that this simple matrix does not correspond to “an equally simple structure of articulatory or acoustic facts” (Fant, 1971, 259). Linell (1973) summarizes the problems involved in a phonological description of Swedish vowels and reviews the most common solutions suggested. The biggest problem is the phoneme /øː/ and its short counterpart /ø/. In articulatory terms /øː/ is a close or near-close and extremely rounded front vowel, while /ø/ is a mid vowel. The articulatory and perceptual distance between the long and short variant is hence large. In the different suggested phonological interpretations, the position of /øː/ and /ø/ has varied in both the height dimension and on the front–back scale. Linell (1973) proposes that /øː/ should be treated as a central vowel phonologically. Other researchers treat /øː/ as a front vowel and use three degrees of roundness (unrounded, out-rounded, in-rounded) for distinguishing /øː/ from /iː/ and /yː/ (Traunmüller & Öhrström, 2007) or from /eː/ and /øː/ (Malmberg, 1956).

A problem with the simple solution in Table 2.2 is that it pushes ò into a more open position than articulatory and acoustic data suggest, and that u is not grouped with the other close vowels, with which it shares important features (see, for example, § 2.3.2.4). Other problems with phonological descriptions of the Swedish vowels have been how to fit in the pre-/r/ variants of ò and ò, and that the long a is a back vowel but it’s shorter counterpart a central vowel. Linell (1973, 12) points out that, interestingly enough, the vowels that present a problem for the phonological description of the Standard Swedish vowels are the same that seem to show considerable variation across Swedish dialects.

\footnote{\(\varepsilon:/\) was not included, because the subjects were asked to pronounce the Swedish letter names. The name of the letter à is pronounced [æ:], while all other letters are pronounced as the corresponding long vowels in Table 2.1.}
Table 2.2. Distinctive features of Swedish vowels. This symmetrical matrix does not correspond to an equally simple structure in articulatory and phonetic terms.

<table>
<thead>
<tr>
<th>front vowels</th>
<th>rounded front vowels</th>
<th>back vowels</th>
</tr>
</thead>
<tbody>
<tr>
<td>close</td>
<td>i(:)</td>
<td>o(:)</td>
</tr>
<tr>
<td>mid</td>
<td>e(ː)</td>
<td>å(:)</td>
</tr>
<tr>
<td>open</td>
<td>å(ː)</td>
<td>ä(:)</td>
</tr>
</tbody>
</table>

2.3.2 Regional variation in vowel pronunciation

The IPA symbols in Table 2.1 are based on Central Standard Swedish. Regional varieties of Standard Swedish show large differences to Central Standard Swedish when it comes to the pronunciation of the vowel phonemes. This concerns especially the long vowels. In the following, the variation in the most variable vowels is described following Elert (2000).

Some regional features affect a number of vowel phonemes in a similar way, while others concern only individual vowels. In §§ 2.3.2.1–4 below features are described that are characteristic for specific geographic regions and affect several vowels in these regional varieties in a similar way. In §§ 2.3.2.5–9, on the other hand, geographic variation in individual vowel phonemes is described.

2.3.2.1 South Swedish diphthongization

In South Swedish spoken language the long vowels are pronounced as rising⁴ diphthongs (Elert, 2000, 38–40). The long front vowels are closing diphthongs that start with a more open pronunciation and end approximately with the Standard Swedish vowel quality. The long back vowels start as unrounded central vowels and move backwards to the Standard Swedish vowel quality. The close vowels are diphthongized more strongly than the open vowels. For example, /iː/ > [ei(j)], /uː/ > [u(u)], /øː/ > [œø] (Elert, 2000, 38).

The amount of diphthongization varies within the South Swedish area, and also depends on how standard-like the speaker talks. The diphthongization is strongest in the (north-eastern part of the) province of Skåne, but is found in the whole South Swedish area.

2.3.2.2 East Central Swedish diphthongization

In Central Sweden, mainly in western and southern parts of Södermanland and south Västmanland, the long vowel phonemes are often pronounced as falling⁴, centering diphthongs (Elert, 2000, 40–42). The vowel quality at the beginning of the vowels corresponds to the Standard Swedish pronunciation, followed by an [e]- or [o]-like

---

⁴The terms “falling” and “rising” are used here for denoting which part of the diphthong that is the prosodically most prominent one. A rising diphthong ends with the more prominent part (e.g., [œø]), while a falling diphthong starts with the more prominent part (e.g., [eː]).
quality. Sometimes laryngalization occurs between these two elements. Examples, /i:/ > [i:(i)\textsuperscript{a}], /u:/ > [u:(\beta)\textsuperscript{a}], /\omega:/ > [\omega:\textsuperscript{a}] (Elert, 2000, 40).

This diphthongization has been shown to be prosodically conditioned (Bleckert, 1987). It is strongest in stressed long vowels and at the end of sentences. It is a rather new feature that was first attested in some Central Swedish towns at the beginning of the 20th century. Strong diphthongization is not commonly accepted as Standard Swedish (Elert, 2000, 42).

### 2.3.2.3 Gotlandic diphthongization

The dialects on Gotland are characterized by both archaic diphthongs and secondary diphthongs (see § 2.3.3). Some of these diphthongs are only part of the rural dialects, but also in standard-like Swedish spoken on Gotland some of the long vowels are diphthongized (Elert, 2000, 42-43). The diphthongization concerns mainly the mid vowels, but the pattern is not as clear as in South Swedish and East Central Swedish diphthongization. Some of the diphthongs are falling, while others are rising; some are closing, while others are centering: /e:/ > [e\textsuperscript{a}], /e:/ > [e:\textsuperscript{a}], /u:/ > [\textsuperscript{a}u] / \omega:/ > [\omega:\textsuperscript{a}], /u:/ > [\textsuperscript{u}u], /\omega:/ > [\omega:\textsuperscript{u}] / (\textsuperscript{a}i:, /\textsuperscript{y}: and /\textsuperscript{a}i: are monophthongs) (Elert, 2000, 42).

### 2.3.2.4 Semi-vowel/fricative ending in long close vowels

The long close vowels are followed by a consonantal segment in large parts of Sweden: /i:/ > [i\textsuperscript{y}], /\textsuperscript{y}:/ > [\textsuperscript{y}u], /u:/ > [u\textsuperscript{b}] and /u:/ > [u\textsuperscript{b}] (Elert, 2000, 43-44). The consonantal ending is most prominent in word final position or before another vowel.

This feature is common in the spoken language in the Central Swedish area. In South Swedish it is less prominent and in Norrland and Finland it hardly exists.

### 2.3.2.5 Damped i and y

The Swedish “damped” i and y are common in the East Central Swedish area (Elert, 2000, 44-45). These variants of the Swedish /i:/ and /\textsuperscript{y}:/ are pronounced with a markedly low F2, so that they sound more retracted than in Standard Swedish, closer to [i] and [u] (Björsten & Engstrand, 1999). They are also co-articulated with a fricative consonant, which gives a characteristic buzzing sound, and, hence, are often transcribed as [\textsuperscript{v}i] and [\textsuperscript{v}y\textsuperscript{f}] (which is not entirely appropriate because the buzzing is present in a larger part of the vowel segment and not only at the end).

The damped pronunciation of i and y has a special distribution. It can be found in some scattered rural dialects, but also in the spoken language in Stockholm and Göteborg. In the areas where the damped pronunciation is part of the rural dialect speakers tend to leave out this feature when they want to speak more standard-like, while in the cities, Stockholm and Göteborg, the damped pronunciation is considered posh (Elert, 2000, 45).
2.3. Swedish vowels

2.3.2.6 Long and short e and ä

In large parts of the language area the short vowels /ɛ/ and /ɪ/ are two separate phonemes (Elert, 2000, 46). However, in East Central Swedish and in the varieties on Gotland and in Finland the two phonemes have merged. Johansson (1982, 94) found that /ɛ/ and /ɪ/ were maintained as two separate phonemes by most speakers in a number of towns in Norrland. For young female speakers, however, she found a perceptual merger of these two phonemes. The merger seems to be more common among younger speakers than among older speakers today, and might be regarded as unmarked standard pronunciation nowadays.

In a smaller geographic area the corresponding long vowels /ɛː/ and /ɪː/ have merged, too, both being pronounced [ɛː]. However, it is not a complete merger, but before /r/ the two phonemes are kept apart. So there is a distinction between, for example, lera [lera] ‘clay’ and lär[a] [lær:a] ‘learn’, but not between leka [leva] ‘play’ and läka [leva] ‘heal’. The merger has been part of the traditional dialects in the surroundings of Stockholm and is also found in Finland. Contrary to the merger of the short e and ä the merger of the corresponding long vowels is not accepted in Standard Swedish pronunciation and there seems to be a reversal of the merger going on (Elert, 2000, 46).

In a study of the language of teenagers in Stockholm, Kotsinas (1994) found three variants of /ɛː/: [ɛː], [ɛː] and [æː]. The pronunciation [ɛː] was more frequent in the lower social class than in the higher social class, while [æː] was more frequent in the higher social class. The open pronunciation [æː] could be interpreted as a reaction against “uncultivated” Stockholm-speech.

2.3.2.7 Long and short u and ö

The pronunciation of /ʊː/ shows considerable variation (Elert, 2000, 47–48). In East Central Sweden a more open pronunciation, [œː], has become popular among younger speakers. Kotsinas (1994) found that the open pronunciation was the most common one among Stockholm teenagers in all social groups. According to Elert (2000, 47), the more open pronunciation is common across generations in some parts of Sweden, for example Östergötland, Västergötland and Värmland.

The vowel written with the letter u in Swedish had the pronunciation [u(ː)] in the 12th and 13th centuries when the Latin alphabet was introduced in Scandinavia. After that the vowel has been fronted leading to [uː] and [oː] in modern Swedish. This process has not had the same speed across the entire language area. Some rural dialects and regional standard varieties have a more central or back pronunciation of u than Standard Swedish, for example, Finland-Swedish and varieties in south-west Skåne, Dalarna and north Västmanland (Elert, 2000, 49).

The short vowels /ɵ/ and /ø/ are merged in the speech of many speakers in East Central Sweden (Elert, 2000, 48). Wenner (2010) showed that younger speakers in Uppland have a shorter acoustic distance between /ɵ/ and /ø/ than older speakers, which suggests an ongoing change. The merger is more common when an /r/ is preceding or following the vowel than in other contexts.
2.3.2.8 The open allophones of ä and ö before r

As mentioned above, the ä and ö vowels have more open allophones that are used only before /r/. However, the degree of openness of these allophones varies. The open allophones are pronounced most open along the east coast and in Finland (Elert, 2000, 48–49). In the west they are more close and in some places there are no allophones, so that the vowels have the same pronunciation in all positions.

In a study of attitudes towards Swedish dialects by Bollek Radovani (2000) not only the pronunciation [ɛː] of /ɛː/ in Uppland, but also the extremely open [æː] before /r/ were seen as stigmatized variants of which the dialect speakers themselves were highly aware.

2.3.2.9 Long and short a

The long a vowel is a slightly rounded open back vowel in Standard Swedish. Among some speakers in Stockholm and Göteborg the vowel is even more rounded and close, almost [ɔː] (Elert, 2000, 50). In Finland the vowel is generally more fronted and unrounded: [a].

In a few areas, the short vowel /a/ has a pronunciation closer to [æ]. Elert (2000, 51) mentions east Småland and Eskilstuna.

2.3.3 Vowel systems of the Swedish dialects

The rural dialects in the Swedish language area differ from Standard Swedish at practically all linguistic levels (segmental, prosody, phonology, morphology, lexicon, semantics, syntax). The vowel systems of the dialects differ from Standard Swedish not only in the pronunciation of the vowel phonemes, but also in the number of phonemes.

The differences in the vowel systems of the Swedish dialects in comparison to Standard Swedish are partly due to archaic features and partly due to innovations. Archaic features are, for example, the preservation of the Proto-Nordic diphthongs /ai/, /au/ and /eu/. These diphthongs were monophthongized in large parts of eastern Scandinavia around 900–1100 and partly merged with original monophthongs. In Standard Swedish these original diphthongs have been replaced by long vowels. The monophthongization spread from the south and never reached some peripheral dialect areas. Dialects on Gotland, in parts of Finland and in parts of Norrland were never affected and have preserved the original diphthongs (Pettersson, 2005).

Another type of diphthongs in the Swedish dialects are the so-called secondary diphthongs. These have developed from original monophthongs which have been diphthongized in some dialects. Secondary diphthongs are characteristic for the southern provinces of Sweden, but are also found on Gotland and in some Finland-Swedish dialects.

Another archaic feature is the preservation of the Proto-Nordic long a. During the 13th and 14th centuries the pronunciation of this phoneme became more close and back resulting in the ä [ɔː] in modern Swedish. This change, too, never reached some
2.4. Acoustic analysis of vowels

A coustic analysis of vowels

Peripheral areas (for example Gotland), which preserved a more open pronunciation [aː]/[^aː].

The Swedish quantity shift, around 1200, affected the vowel systems of the dialects in different ways. In Proto-Nordic four types of syllables could occur in stressed positions: VC, V:C, VC: and V:C:. The quantity shift resulted in the loss of VC and V:C: sequences in stressed syllables. In modern Standard Swedish and most dialects all stressed syllables have either a long vowel followed by a short consonant (V:C) or a short vowel followed by a long consonant or a consonant cluster (VC: or VCC). During the quantity shift, the new system was reached in different ways by different dialects. In some dialects the vowel in VC sequences was lengthened, while in other dialects the consonant was lengthened. Also the quality of the short vowels was affected differently in different dialects (Pettersson, 2005, 224–227). Some dialects were not affected at all, or only partly, by the quantity shift (mainly dialects in Norrland and Finland).

Even though there is a long tradition of dialect research in Sweden and Finland, the phoneme systems of Swedish dialects as spoken nowadays are largely unknown. Structural descriptions of the vowel systems of dialects are rare, and because of the large changes that the dialects have gone through since the middle of the 20th century older descriptions of the phoneme systems are not necessarily valid anymore. Many dialects have not been the subject of recent studies, and even for dialects that have recently been studied, describing the phoneme system is not always easy. In a phonological analysis of the Swedish dialects of the province Österbotten in Finland, known for well-preserved rural dialects, Wiik (2002) concluded that especially the long vowel systems are difficult to describe. The traditional rural dialects have more long vowels than short vowels, mainly due to diphthongs among the long vowels. However, because of dialect leveling the long vowel systems are unstable. There are, for example, tendencies to monophthongize diphthongs, and in words borrowed from Standard Swedish phonemes that have not originally been part of the rural dialect, are used. The study by Wiik (2002) included the dialects of 30 parishes in Österbotten. Only a few of these dialects showed identical long vowel systems. The number of long vowel phonemes varied between ten and sixteen, while the number of short vowel phonemes was less variable (between seven and nine short vowel phonemes).

2.4 Acoustic analysis of vowels

When studying phonetic differences between languages or language varieties we can choose to work either with phonetically transcribed data or with acoustic analysis of speech samples. A disadvantage of phonetic transcription is that it requires a large amount of work and still is not a very exact method. Linguistic experience will influence the decisions even of highly trained transcribers (Dioubina & Pfitzinger, 2002), and transcribers do not even always agree with themselves when transcribing the same utterance multiple times (Pfitzinger, 2003).
Numerous efforts have been made to assess phonetic quality of speech segments directly from the acoustic signal. The most common way of measuring vowel quality acoustically are formant measurements, explained in § 2.4.1. Another approach are so-called whole-spectrum methods. Different whole-spectrum methods are discussed in § 2.4.2.

Analyzing vowel quality acoustically is a more objective method than using transcriptions, but does not come without its own problems. The acoustic signal includes information that is not linguistically meaningful, but has to do with anatomical/physiological differences between speakers (§ 2.4.3). Transcribers are very good at ignoring spectral differences that are due to speaker-specific variation and report the actual linguistic differences that they perceive. But when working with acoustic speech data, normalizing for the non-linguistic speaker-dependent variation has always been problematic. Different approaches to speaker normalization are discussed in § 2.4.4.

2.4.1 Measuring formants

Since the work of Peterson & Barney (1952) formant measurements have been the classical way of measuring vowel quality. The two first formants (i.e. the lowest resonance frequencies of the vocal tract during pronunciation) are the most distinctive acoustic parameters that determine vowel quality. These correspond relatively well to the articulatory vowel features height and advancement. This can be seen in the spectrograms in Figure 2.3. The horizontal axis shows time, while the vertical axis shows frequency. The intensity of the speech signal at different frequencies at a given time is expressed by the darkness of the shading. The formants show up as darker bands in the spectrogram. [i] and [u] which are both close vowels have a low first formant (F1, the lowest formant), while the open vowels [æ] and [a] have a higher F1. The front vowel [i] has a high frequency for F2, while the back vowel [u] has a very low F2. The F2 values for [æ] and [a] are in between these two extremes.

A two-dimensional graph, Figure 2.4, where the two dimensions represent tongue advancement and height has been used by phoneticians to display the main parameters of vowel production. Joos (1948) showed that vowels plotted in an acoustic plane, where F2 is represented on the horizontal axis and F1 on the vertical axis with the origin in the upper right corner, leads to a relatively similar configuration as the articulatory graph. This parallel between the articulatory and acoustic planes is indicated with the arrows in Figure 2.4. However, the correspondence between formants and vowel height and advancement is not perfect. Rosner & Pickering (1994, 46–48) show that altering the three articulatory parameters constriction size, constriction location and mouth opening in vowels does not have any one-to-one correspondence with formants, but both the F1 and the F2 values are influenced by the change in any of the articulatory parameters.

The third formant also plays a role in vowel categorization, but this is less well understood than F1 and F2. The importance of F3 seems to vary across languages, and for some languages the first two formants are enough for listeners to identify all
2.4. Acoustic analysis of vowels

Figure 2.3. Vowel spectrograms of [i], [æ], [a] and [u]. The dashed lines mark the approximate positions of the formants.

Figure 2.4. Vowel quadrilateral (vowel chart of the International Phonetic Association) with formant affiliation. F1 and F2 (Hz) with values of a typical male voice.
vowels (for a discussion of this see Adank, 2003, 38–39). However, phoneticians warn against using only F1 and F2 in a routine manner since some valuable data present in third formant might get lost. Fujimura (1967) found that for the identification of Swedish front rounded vowels F3 seems to be important.

In Figure 2.3 the three first formants can be most clearly distinguished in the case of [æ]. The vowel [i] has a large distance between F1 and F2, while F2 and F3 are so close to each other that they are hard to distinguish. In the cases of [a] and [u] F1 and F2 are very close to each other.

Linear predictive coding is the most commonly used computational method for estimating formant frequencies from a speech signal (Rosner & Pickering, 1994, 8–11). A problem with algorithms for automatically determining formant frequencies in a spectrum is that they always give some false measurements. Formants very close to each other (like F1 and F2 in [u]) sometimes cannot be separated by the algorithms, and on the other hand the algorithms sometimes find false formants in the gap between formants if two adjacent formants have a large distance (like F1 and F2 in [i]). In a study of Swedish vowels by Eklund & Traunmüller (1997), where formants were first measured automatically and subsequently checked manually, corrections had to be made for approximately a quarter of the vowel segments. Likewise, Adank, Van Hout, & Smits (2004) report that the automatic formant measurements of 20–25% of the vowels in a Dutch study had to be corrected manually.

Formant measurement has been used for measuring vowel quality in a number of large-scale studies of regiolects and sociolects. Labov, Ash, & Boberg (2005) described regional varieties of North American English by measuring the first two formants of on average 300 vowel tokens of 439 speakers. Adank, Van Hout, & Van de Velde (2007) investigated regional differences in vowel pronunciation in Standard Dutch based on measurements of duration and formant frequencies of 15 vowels of 160 Dutch and Flemish speakers. Clopper & Paolillo (2006) analyzed formant frequencies and duration of 14 American English vowels as produced by 48 speakers from six dialect regions. However, the need for manual correction of the data makes formant measurements a very time-consuming method when working with data sets including a large number of speakers.

2.4.2 Whole-spectrum methods

Besides formant-based approaches to measuring vowel quality whole-spectrum methods are also used. These are sometimes preferred since they include more acoustic information than formant frequencies. Moreover, whole-spectrum methods can generally be more reliably automated than formant analysis, which makes them more suitable for fast analysis of large amounts of data.

In the 1960s, Dutch researchers introduced principal component analysis (PCA) of band-pass-filtered\(^5\) spectra as a method for identifying vowels acoustically (Plomp, 1967). Applying an acoustic filter means that sound energy above and/or below a given frequency in
2.4. Acoustic analysis of vowels

Pols, & Van de Geer, 1967). Vowel spectra were band-pass filtered with $\frac{1}{3}$-octave filters, roughly corresponding to the ear's critical bandwidth. The vowel spectra were filtered up to 10,000 Hz, which, after combining the lowest filters in order to reduce the effect of fundamental frequency and for even better correspondence with the critical bandwidth, resulted in each vowel being described by the sound-pressure levels in decibels in 18 pass bands. PCA was used to find the linear combinations of the pass bands in the filter bank representation that would explain most of the variance in the data. The two first principal components (PCs) turned out to explain most of the variance (68.4%), while extracting up to four PCs significantly improved the amount of explained variance (84.1%). The PCs were shown to correlate highly with perceptual dimensions used by listeners to identify vowels (Pols, Van der Kamp, & Plomp, 1969). The method was mainly seen to be a useful application within automatic speech recognition. An advantage of a filter bank representation over measuring formants is that using band-pass filtering and PCA is much faster than formant analysis and does not require manual correction. The two first PCs correspond well with the two first formants. Correlating average formants values of twelve vowels pronounced by 50 male speakers showed very high correlation ($r = 0.989$ and $r = 0.993$) with the two first PCs after an optimal rotation of the PC plane (Pols, Tromp, & Plomp, 1973). However, correlating formants with the PC values of the individual speakers showed significantly less correspondence. In an automatic recognition task vowels were identified about equally well using formants and PCs (Pols et al., 1973).

Van Nierop, Pols, & Plomp (1973) compared the results of a PCA on band-pass filtered spectra of twelve Dutch vowels by 25 female speakers to the earlier results by Klein, Plomp, & Pols (1970) and Pols et al. (1973) based on 50 male speakers pronouncing the same vowels. The eigenvectors of the two first components were rather similar for the female and male data, but the peaks in the curves of the female analysis occurred about $\frac{1}{3}$-octave higher than the corresponding curves for the male data. The extracted components were not identical for the female and male data, but a small rotation of one of the two configurations showed that there was almost a complete convergence of the two solutions after rotation.

Jacobi (2009) applied PCA to band-pass filtered spectra in a study of variation in Dutch diphthongs and long vowels (/ei/, /ou/, /oe/, /er/ and /e:/) among 70 speakers (35 female and 35 male). Her aim was to find out if the sub-phonemic variation in these vowels could be related to socio-economic status and the age of the speakers. Band-pass filtering was chosen over formant analysis, since it can be fully automated without the need for manual correction of errors and yet, through high correlation with formants, offers interpretability in terms of articulatory and perceptual attributes. However, applying the method introduced by Plomp et al. (1967) to a variational linguistic study presented some challenges. The experiments of the Dutch researchers in the 1960s and 1970s were done in controlled environments.

---

[1] A sound spectrum is filtered away. The part of the spectrum remaining is called a pass band. A band-pass filter can be characterized by its center frequency and bandwidth. See Johnson (2003, 14-17).
Jacobi’s data, on the other hand, consisted of spontaneous speech from the *Corpus Gesproken Nederlands* (= Corpus of Spoken Dutch) in extremely varying recording situations. Plotting the speakers /a/-/i/-/u/ vowel triangles in the PC2/PC1 plane showed a remarkable variability between speakers in the positioning of the triangles. The differences could be traced back to the signal-to-noise ratios in the recordings, which varied from interviews in silent environments to private conversations and broadcast recordings with music in the background. It turned out that the lower the signal-to-noise ratio (that is, the more noise), the smaller the measured vowel space in the PC plane and the higher the PC values. Vowels produced by men and women were analyzed together in the PCA. Some significant differences in the PC values between the sexes were found, but these were small compared to the differences depending on the varying recording situations. In order to normalize for the differences in position and size of the vowel spaces of different speakers, relative PC values were used: the positions of all vowels in the PC plane were related to the speaker’s point vowels /i/ and /a/.\(^6\) In order to enhance the interpretability of the PCA solution Jacobi used only point vowels to build up the PCA. When an equal number of all point vowels are used in the analysis phase of the PCA, all articulatory-acoustic dimensions are represented equally and all possible differences in vowel quality are accounted for. The results can then be used to represent all other vowels within the vowel space.

For the band-pass filtering, Jacobi used Bark filters instead of \(\frac{1}{3}\)-octave filters. The Bark scale is a psycho-acoustical scale corresponding to the critical bandwidth of human hearing, which means that a representation in Bark filters should correspond to how humans perceive vowel sounds. The Bark scale is roughly linear up to 1000 Hz and roughly logarithmic at higher frequencies. Jacobi found high correlations between PCs and formants (see § 5.2.1) and concluded: “pc1 and pc2 of a PCA on barkfiltered /a/, /i/, /u/ yielded comparable results to F1Bark and F2Bark, and are thus easily interpretable in terms of articulation” (Jacobi, 2009, 42).

In automatic speech recognition Mel frequency cepstral coefficients (MFCCs) are widely used today for identifying speech sounds. Like the Bark scale, the Mel scale is an auditory scale (a doubling of Mels corresponds to the doubling of perceived pitch). MFCCs are obtained by applying discrete cosine transformation to Mel-scaled bandpass filters (Harrington, 2010). In a study exploring cluster analysis as a method for classifying dialects Lundberg (2005) applied MFCCs to vowel data from 285 subjects from 95 sites in the Swedish language area (see also § 2.2.3). A reduction of twelve MFCCs to two dimensions by multidimensional scaling showed a configuration similar to the IPA vowel quadrilateral (Lundberg, 2005, 26).

For the present study PCA on Bark filtered vowel spectra was chosen for analyzing vowel segments acoustically. The main reason for choosing a whole-spectrum method instead of measuring formants is that the whole-spectrum methods have been shown to be more reliably automatable than formant measurements. Choosing a method which does not need manual correction was important, because the data

\(^6\)Because /u/ was the most sensitive point vowel to the signal-to-noise ratio only /i/ and /a/ were used for calculating the relative positions.
2.4. Acoustic analysis of vowels

for the present study includes nearly 1,200 speakers (see Chapter 4). MFCCs could have been chosen instead of Bark filters and PCA, and might yield similar results.

The method and its application to the present data set is described more closely in Chapter 5. In addition to band-pass filtering, formant analysis was applied to a smaller subset of the data in order to validate the method and to find the optimal PCA configuration (see § 5.2).

2.4.3 Speaker-dependent variation in vowels

A challenge for all acoustic methods is the speaker-dependent differences in speech sounds due to the different size and shape of the vocal tract of speakers. Two vowels uttered by two different speakers can have very similar formant frequencies while they are perceived as belonging to different phonemes by listeners of the languages. And, conversely, two vowels sounding the same to listeners might differ in formant frequencies. Generally a smaller vocal tract generates higher formant frequencies. This is why children have higher formant frequencies than adults and women generally higher than men (Peterson & Barney, 1952). But also within age and gender groups individuals show a lot of variation.

Figure 2.5 shows an example of the inter-speaker variability in formant measurements. In the graphs the formant frequencies in Bark of ten Swedish long vowels produced by three adult speakers, two female (yw_1 and yw_3) and one male (ym_2), are plotted. The speakers all speak the same dialect, the dialect of Malung. The data come from the data set discussed in § 5.2 below. It is evident that the vowel spaces vary both in size and position in the F2/F1 plane (scales are reversed in order to resemble the articulatory vowel quadrilateral). If looking at the absolute formant frequencies in Bark, the [u:] in sot of speaker yw_1 is closest to the [o:] in läs of ym_3. The [æ:] (lär) of the male speaker (ym_2) is closest to the [œ:] (dör) of the two female speakers. However, the relative positions of the vowels in all speakers’ vowel spaces resemble each other a lot.

Inter-speaker differences, like the ones present in formant measurements, are usually found also when using whole-spectrum methods for vowel analysis.

The mechanisms behind the speaker normalization that listeners perform are not fully understood. The problem has been viewed both from a vowel-extrinsic and from a vowel-intrinsic perspective. Extrinsic adaptation means that listeners adapt to every speaker’s vowel system as a whole, while intrinsic adaptation means that the information needed for normalization is included in each segment itself. Intrinsic factors have been sought in distances between fundamental frequency and higher formants. Fundamental frequency (F0) seems to be an important cue for identifying vowels in mixed speaker conditions as shown in experiments by Nusbaum & Morin (1992), Eklund & Traunmüller (1997) and Halberstam & Raphael (2004).
Figure 2.5. The vowel spaces of three speakers of the same dialect (Malung). Ten different long vowels are plotted in the F1–F2 plane. The words used for eliciting the vowels are displayed: dis [iː], dör [œː], lat [aː], leta [eː], lus [uː], lås [oː], lär [æː], sot [ʊ], söt [œː], typ [yː].
2.4. Acoustic analysis of vowels

Table 2.3. Division of formant-based speaker normalization procedures into four groups. The division was introduced by Adank (2003).

<table>
<thead>
<tr>
<th>Vowel-intrinsic</th>
<th>Formant-intrinsic</th>
<th>Formant-extrinsic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>scale transformations (e.g. Bark, Mel)</td>
<td>Syrdal &amp; Gopal (1986)</td>
</tr>
<tr>
<td>Vowel-extrinsic</td>
<td>Gerstman (1968)</td>
<td>Nordström &amp; Lindblom (1975)</td>
</tr>
<tr>
<td></td>
<td>Neary (1977)</td>
<td></td>
</tr>
</tbody>
</table>

2.4.4 Speaker normalization

A wide range of different normalization procedures for vowels have been suggested in different studies. The choice of a normalization procedure depends crucially on the aim of the normalization. For example, in automatic speech recognition systems the aim is to remove all other kinds of variation but the phonemic. For sociolinguistic and dialectological research on the other hand we are interested in also maintaining sub-phonemic variation. The aim of speaker normalization procedures has been described by Disner (1980) as follows: “they should maximally reduce the variance within each group of vowels presumed to represent the same target when spoken by different speakers, while maintaining the separation between such groups of vowels presumed to represent different targets”.

2.4.4.1 Formant-based normalization procedures

Adank (2003) divided formant-based normalization procedures into formant-intrinsic and formant-extrinsic, as well as vowel-intrinsic and vowel-extrinsic (Table 2.3). Vowel-intrinsic means that the normalization can be done using information present in a single vowel token, while vowel-extrinsic means that the whole vowel space of a speaker, or at least the point vowels, are taken into account when performing the normalization. Formant-intrinsic means that each formant can be normalized without knowledge of higher or lower formants (including F0). Formant-extrinsic methods use information across formants, for example, relative distances between formants (including F0).

A vowel-intrinsic/formant-intrinsic method is basically a transformation of formant values according to some scale. The Hertz scale that is used for measuring frequencies is a linear scale. Bark and Mel are examples of auditory scales, which correspond better to human perception, which is roughly linear up to 1000 Hz and roughly logarithmic at higher frequencies. Transforming the measured Hertz values to an auditory scale improves the correspondence between acoustic measures and perception.

A vowel-intrinsic/formant-extrinsic procedure was used by Syrdal & Gopal (1986). In this method the data was transformed by measuring the distances in Bark between adjacent formants.

Several methods use information across vowel categories for speaker-normalization (vowel-extrinsic methods). Gerstman (1968) introduced a method for normaliz-
ing on the basis of the highest and lowest F1 and F2 values of each speaker. Lobanov (1971) did a z-transformation of the F1 and F2 values of every speaker. Neary (1977) applied a log transformation to the Hertz frequencies and centered the formant values of every speaker. In Neary’s approach the center of every speaker’s vowel spaces was moved to zero, but varying sizes of the vowel spaces of different speakers were not normalized for. Labov (1994) has used Neary’s log mean transformation in his sociolinguistic studies.

A problem with the vowel-extrinsic methods is that they assume that either the average values or the extreme values of all speakers are linguistically stable. The approaches of Lobanov (1971) and Neary (1977) imply that all speakers use the same phoneme system and that data of all vowel categories are present for all speakers. Otherwise the average values of different speakers would not be comparable. In the procedure by Gerstman (1968) a reliable transformation is possible only if the vowel phonemes representing the highest and lowest F1 and F2 values are linguistically stable without sub-phonemic variation.

The vowel-extrinsic/formant-extrinsic normalization procedures use information across vowel tokens and across formants. In addition to being vowel-extrinsic/formant-extrinsic the normalization procedure by Nordström & Lindblom (1975) is speaker-extrinsic. In their study Nordström & Lindblom (1975) started by estimating the average length of vocal tracts of men and women by calculating average F3 values of open vowels within each sex group. Because the length of the vocal tract correlates highly with formant frequencies, a scale factor based on the ratios between average male and female vocal tract lengths (as estimated by F3) could be used for transforming the formant frequencies. This procedure does not remove speaker-dependent differences within the two sex groups, but it normalizes for the systematic differences between male and female voices.

The normalization procedure of Miller (1989) is based on distances between formants, hence, formant-extrinsic. Furthermore a speaker-specific anchor point based on average F0 values is used, which makes it vowel-extrinsic.

### 2.4.4.2 Normalization in whole-spectrum approaches

A centering of data, corresponding to the procedure of Neary (1977), was used on bandpass-filtered vowel data by Pols et al. (1973). Jacobi (2009) applied centering to principal components of bandpass-filtered data, but decided on using another method for speaker normalization. The point vowels /i/ and /a/ were considered stable point vowels in the Dutch data in Jacobi’s study. Hence, positions relative to each speaker’s /i/ and /a/ were calculated for all vowels, which accounted for differences in the size and position of the vowel spaces of different speakers.

Both the method of Pols et al. (1973) and the one of Jacobi (2009) are vowel-extrinsic and, therefore, depend either on identical phoneme systems or stable point vowels in the data.
2.4.4.3 Evaluations of normalization procedures

When comparing different vowel normalization procedures Disner (1980) found that Neary’s log-mean procedure was best at reducing scatter within languages. But a crucial result from her study was that most of the procedures tested performed quite poorly when comparing vowels from different languages, and some procedures even reversed the linguistic trends. Procedures like those of Neary (1977) and Lobanov (1971), which use mean values and/or standard deviations of formants for normalizing, fail in comparing language varieties with different phoneme systems because the means and standard deviations of these systems are not comparable. These procedures are good at indicating the relative position of vowels within the phonetic space of one language, but the relative positions do not represent the positions in a universal phonetic space.

One of the few procedures that Disner (1980) found to be valid for cross-language comparison was the PARAFAC procedure of Harshman (1970), which can be used for isolating and then averaging speaker-dependent differences. A problem with PARAFAC, however, is that it assumes a priori knowledge of the phoneme category of each vowel token. Hence, it cannot be used for data sets where manual categorization is not done prior to the acoustic normalization.

Adank (2003) evaluated a number of speaker normalization procedures, among others, the ones in Table 2.3 (p. 31), with the criterion that a successful normalization procedure should preserve phonemic variation and sociolinguistic speaker-related variation, but minimize the anatomical/physiological speaker-dependent variation. In within-language comparison, that is, when comparing varieties with the same phoneme system, the vowel-extrinsic/formant-intrinsic normalization procedures perform quite well. Adank (2003) found that Lobanov’s (1971) procedure was most effective in reducing speaker-specific variation while maintaining sociolinguistic variation. The vowel-intrinsic procedures performed poorly in Adank’s evaluation and the formant-extrinsic procedures performed worse than the formant-intrinsic ones. Adank’s results support the results of Disner (1980), and Adank (2003, 183) strongly emphasizes that “it is not advisable to carry out normalization procedures on data sets that are not fully phonologically comparable”.

In studies of formant frequencies that include languages or dialects with different phoneme systems or deviating vowel spaces, speaker normalization remains a problem. A common solution when wanting to compare formant values of different varieties is to use raw formant measurements, but to average over a number of speakers (for example, Adank et al., 2007). Averaging over a sufficient number of speakers will reduce the effect of varying sizes of vowel spaces of individual speakers. However, this method can be used only when one can assume that there are no systematic differences in the size and shape of the vocal tracts across the different groups of speakers. Female and male speakers cannot be directly compared to each other by using within-group averages, because of the systematically lower formant frequencies of men than women. Neither was using group averages an option in a study by Yang (1996) where American English and Korean vowels were compared,
because there was a considerable difference in vocal tract length between American English and Korean speakers. In order to be able to normalize for the differences in vocal tract length between the speaker groups, Yang (1996) estimated the vocal tract length by applying the method of Nordström & Lindblom (1975), originally developed for normalizing for the differences between male and female voices (see § 2.4.4.1).

The varieties analyzed in the present study are not fully phonologically comparable, which makes it impossible to use most of the normalization procedures mentioned above. Neither can stable point vowels be found in the data, which excludes the possibility to use a normalization procedure comparable to the one used by Jacobi (2009). Using raw acoustic measurements and averaging over groups of speakers is complicated by the small number of speakers per variety. At each site, around twelve speakers were recorded representing both sexes and two generations of speakers (see § 4.3). The number of male and female speakers is not constant across all sites and vowels, and considerable variation across the two age groups can be expected (see § 2.1) making a comparison of the two age groups necessary, which reduces the number of speakers per variety.

For these reasons, a solution comparable to the one by Nordström & Lindblom (1975) was sought. That is, the aim was not to try to remove all speaker-dependent variation, which would be virtually impossible, but to normalize for the systematic differences between male and female voices. The systematic differences in vowel spectra produced by men and women found by Van Nierop et al. (1973) (see § 2.4.2) was used as a basis for the normalization. The procedure is described in Chapter 5.

2.5 Dialect geography and dialectometry

In dialectology, regional differences in language use are studied. Dialect geography, more precisely, maps geographic distributions of dialectal features. The first large-scale dialect geographic project was the one by George Wenker in Germany in 1876. Questionnaires were completed by about 45,000 schoolmasters in Germany, and based on the survey Wenker drew maps by hand, each map representing a dialectal feature. This project was followed by many dialect atlas projects in Europe and North America in the first half of the 20th century. The usual way to present the data in dialect atlases was creating display maps representing one single feature. In secondary studies based on data from dialect atlases interpretive maps, grouping variants into the most predominant groups, are also found.

The traditional method of identifying dialect areas has been the isogloss method. The term isogloss was introduced by J. G. A. Bielenstein in 1892 (Chambers & Trudgill, 1998, 89). An isogloss is a line on a map that is drawn between locations where speakers use different variants of a feature. When many isoglosses are drawn on the same map some patterns can usually be identified. When many isoglosses coincide they form an isogloss bundle, which usually indicates a major dialect division. A well known isogloss bundle from Swedish dialectology is the one distinguishing
South Swedish dialects (Andersson, 2007, 42). South from the isogloss bundle a dorsal /r/ is used, while apical /r/ is predominantly used in other Swedish varieties. Other isoglosses coinciding closely with the /r/ isogloss are the lack/use of retroflex consonants and the use of a retroflex flap for /rd/.

The problem of identifying dialect areas by means of isoglosses has been formulated by Chambers & Trudgill (1998, 96–97) as follows:

“It is undeniable that some isoglosses are of greater significance than others, in the sense that some mark distinctions ‘felt’ to be culturally important while others do not, some persist while others are transitory, and the like. It is equally obvious that some bundles are more significant than others, in the same sense. Yet in the entire history of dialectology, no one has succeeded in devising a satisfactory procedure or a set of principles to determine which isoglosses or which bundles should outrank some others. The lack of theory or even a heuristic that would make this possible constitutes a notable weakness in dialect geography.”

In the middle of the 20th century dialect geography as an international discipline declined. One of the reasons was that proper methods for analyzing the huge data collections in the dialect atlases were lacking (Chambers & Trudgill, 1998, 20). A revitalization started in the 1980s when computational methods offered new possibilities for analyzing the data. By this time sociolinguistics had also developed a new theoretical framework for analyzing linguistic variation.

When using isoglosses for dividing language areas into dialect regions, the choice of which linguistic features to emphasize is subjective and different researchers are likely to make different choices. As a more objective alternative to the isogloss methods Séguy (1973) introduced dialectometry (Fr. dialectométrie). Séguy worked with data from Atlas linguistic de la Gascogne. Instead of only drawing isoglosses based on single features, Séguy started by calculating dissimilarity scores between dialects based on all available linguistic features in the atlas. The linguistic distances between adjacent dialects were plotted on maps using thicker or thinner lines indicating, respectively, larger or smaller distances. The main dialect borders could be identified by the thickest lines. At the same time the maps showed that the dialect landscape is continuous without many abrupt borders.

Goebi (1982, 1984, 2006) adapted Séguy’s ideas and started using computational methods for calculating linguistic distances between varieties. Goebi extended the dialectometric idea by not only calculating the similarity between geographically adjacent dialects, but between all varieties in a data set. The result is a \( n \times n \) similarity matrix, where \( n \) is equal to the number of dialects in the data set. Like Séguy’s data, Goebi’s data originates from dialect atlases (Atlas linguistique de la France and Atlante italo-swizzero). The similarity between two dialects is calculated by counting for how many features in the data set two dialects use the same variant and for how many features they differ. The percentage of similar variants is the similarity measure. In Goebi’s method the linguistic similarity is plotted on maps by choosing one reference site and by a color scheme displaying the similarity with
all other sites (thus using only one vector with \( n \) values from the \( n \times n \) similarity matrix). With this approach \( n \) different maps can be created of a data set with \( n \) sites. Each map shows how the linguistic landscape looks from the view of the speakers of one specific dialect. A problem with this approach is that in a data set with a large number of data sites a large number of maps can be created, which complicates interpretation.

Goebel also introduced cluster analysis as a dialectometric method. In cluster analysis the distances between all dialects are analyzed and a partitioning is made of the data so that the most similar dialects will belong to the same group. With cluster analysis dialect areas can be detected and the distances are not observed from one specific site, but the perspective is that of an “objective observer”.

Kessler (1995) introduced the Levenshtein distance (also called string edit distance) as a tool for measuring dialect distances. This is a more refined measure than the binary counts of Séguy and Goebel, since phonetic similarity of segments can be taken into account (for a detailed description, see Nerbonne & Heeringa, 2010). Gooskens & Heeringa (2004) validated phonetic distances between Norwegian dialects measured with the Levenshtein distance with perceptual distances and found a high correlation. Other methods for aligning transcriptions and measuring dialect distances have been proposed as well, for example, Pair Hidden Markov Models (Wieling, Leinonen, & Nerbonne, 2007) and an iterative pairwise aligning algorithm used to produce multiple sequence alignments (Prokip, Wieling, & Nerbonne, 2009).

Heeringa (2004) applied the Levenshtein distance to Dutch and Norwegian dialect data. In addition to applying cluster analysis to the \( n \times n \) distance matrices with the aggregate linguistic distances between all varieties, Heeringa used multidimensional scaling (MDS). MDS is a technique for visualizing pair-wise distances in a low dimensional space. Based on the pair-wise distances, positions in a low-dimensional space that approximate the original distances can be calculated (see also § 7.1 below). While Goebel’s method for visualizing distances between dialects allows visualization of distances only from one reference point at a time, the advantage of MDS is that the relative linguistic distances between all varieties are displayed simultaneously. MDS has been used in linguistics since Black (1973). The results of MDS are usually visualized in two-dimensional Cartesian coordinate systems. In the coordinate system the positions of lects reflect the linguistic distances instead of showing geographic distances like a map. Chambers & Trudgill (1998, 147) comment on the use of MDS in dialectology:

“One possible objection to multidimensional scaling is that it eliminates geographic distance in favor of statistical distance, so to speak. However, it is the difference between the two types of distances that proves to be one of the most telling aspects of the analysis.”

When MDS is applied to linguistic data three dimensions usually explain so much of the total variance in the data that scaling to more than three dimensions is not considered necessary. Wilbert Heeringa and Peter Kleiweg\(^7\) found a way to map
2.5. *Dialect geography and dialectometry*

the results of three-dimensional MDS to geography by using the RGB color model (Nerbonne, forthcoming). The method is explained more closely in Appendix B of this thesis. By using the three-dimensional color scheme for coloring maps, a three-dimensional linguistic space is linked to the two-dimensional geographic space. This facilitates the interpretation of results obtained by MDS. The method is used for visualizing dialect distances throughout Chapter 7 in the present thesis.

In the dialectometric work by Heeringa and Nerbonne (e.g., Heeringa, 2004; Nerbonne & Siedle, 2005) cluster analysis and MDS have been used side by side. While cluster analysis detects dialect areas, the results of MDS show that the dialectal variation within the areas and at the boundaries is actually continuous and that the borders are not as abrupt as suggested by clustering. These complementary analyses are in line with how traditional scholars have dealt with geographic variation. At the same time as researchers agree that linguistic variation is gradual, they have found it important to group dialects into larger areas and describe which the most prominent dialectal features are for the distinguished areas.

Recent research has shown that cluster analysis should be applied with caution to dialect data (Nerbonne, Kleiweg, Heeringa, & Manni, 2008; Prokić & Nerbonne, 2008). Small differences in the input data can lead to substantially different clustering results. Because of this, the results of different clustering algorithms should be compared and the results should be carefully evaluated. In data that is truly continuous clustering algorithms are unlikely to find meaningful clusters.

A different approach to detecting dialect areas was proposed by Hyvönen, Leino, & Salmensivuori (2007) and Leino & Hyvönen (2008). They worked with data very different from the data in the previously mentioned studies. The data comprised lexical distribution maps from the *Dictionary of Finnish Dialects* and was binary: either a lexical item had been recorded at a municipality or it had not. The data suffered from uneven sampling; some municipalities had been thoroughly sampled, while the data from other places was sparse. Hence, the absence of a record of a lexical item at a site meant either that it was not used in that dialect or that it just did not happen to be recorded. While cluster analysis and MDS did not perform well on the data, different component models (factor analysis, non-negative matrix factorization, aspect Bernoulli, independent component analysis and principal component analysis) detected distribution patterns corresponding to dialect areas. With these methods it was also possible to factor out the effect of uneven sampling. These methods do not make sharp divisions into dialect areas like cluster analysis, but show core areas and transitional zones. The five different component methods compared by Leino & Hyvönen (2008) all presented the data in slightly different ways. A conclusion of the study was that factor analysis was the most stable method producing the most easily interpretable results.

The dialectometric research tradition which started with Séguy and was continued by, among others, Goebel, Heeringa and Nerbonne, has focused on the aggregate analysis, that is, the picture that emerges when all available data is considered (Ner-
bonne, 2009). This was a reaction against the isogloss method, which analyzes only one variable at a time or a limited number of isoglosses at best. However, a drawback of the aggregate measures of linguistic similarity/dissimilarity used in dialectometry is that it is hard to trace back the linguistic features characterizing the linguistic areas that have been detected in the aggregate. Moreover, the aggregate distances can hide varying distributions of dialectal features in the original data.

Sometimes different distribution patterns can reveal more about the causes of dialectal variation than an aggregate analysis does. For example, in Scandinavian dialectology two centers have traditionally been identified from which novel features have spread: a southern and a central one. Features that have spread from the south have a different distribution than features that have spread from the center, and the influence of these two centers can be traced back to different time periods (Pettersson, 2005). While concentrating on the aggregate analysis, dialectometric methods are likely to ignore the different underlying distribution patterns below the aggregate level, like the two different innovation centers in Scandinavia. Aggregate analysis gives a view of the relationships between dialects, but in order to explain the relationships the diffusion patterns are important.

Some attempts have been made to identify linguistic structure in the aggregate analysis. Prokić (2006) applied aggregate dialectometric analysis to Bulgarian dialect data and, in addition, extracted the most frequent regular sound correspondences between dialects from the same data set. She found that out of the ten most common regular correspondences eight were correspondences between two vowels or insertions/deletions of vowels, which suggests that the vowels are largely responsible for the classification attained by the aggregate analysis. For each of the regular sound correspondences a map could be created showing the geographic distribution.

Nerbonne (2006) applied factor analysis to vowel data from the Linguistic Atlas of the Middle and South Atlantic States, which contains transcribed lexical items. Vowels were identified by the word they were extracted from (for example, the first vowel in the word “afternoon”). Factor analysis revealed which vowels could explain most of the variance and also which vowels had similar distributions. However, manual investigation of the data was needed in order to identify the most important variants of the vowels involved.

Factor analysis was also used by Clopper & Paolillo (2006) to analyze formant frequencies and duration of 14 American English vowels as produced by 48 speakers from six dialect regions. The analysis showed regional patterns and co-occurrence of some vowel features, but the analysis was complicated by interactions with speaker-sex.

While the essential difference between dialectometry and traditional dialect geography from the start has been the focus on aggregate analysis in dialectometry, the word dialectometry literally means ‘measuring dialect’. Literally dialectometry could, thus, include any quantitative/computational analysis of dialects. Methods first introduced in dialectometry have also spread to research that does not include an aggregate approach (for example, studies mentioned in § 2.2.3). As large digitalized speech corpora become available for analysis, the benefits from using statist-
2.5. **Dialect geography and dialectometry**

Typical, multivariate methods become obvious. These methods can find relationships in data sets which are too large and complex for manual analysis. Besides being able to handle large amounts of data, data-driven methods also make the analysis more objective. Often patterns are recognized that were not expected by the researcher (Chambers & Trudgill, 1998, 141). Nonetheless, the analysis is of course only as objective as the input, and the data for statistical analysis should be carefully chosen and interpreted. It should also be noted that the methods used in dialectometry are usually exploratory and not confirmatory. They can be used to describe data, but not to test hypotheses.

In the present thesis, factor analysis (FA) and multidimensional scaling (MDS) were applied in order to explore the dialectal variation in Swedish vowel pronunciation. Leinonen (2008) showed that, in contrast to cluster analysis and MDS, FA is able to detect different diffusion patterns in dialect data and find co-occurring features. In the paper mentioned, vowel quality in Swedish dialects was measured at the temporal midpoint of each vowel segment and only geographic variation was analyzed. As described in § 2.3, diphthongization is an important characteristic for regional varieties of Swedish. Therefore, the present study extends the analysis of Leinonen (2008) by adding spectral change as a variable in FA. Moreover, not only geographic variation, but also within-site variation is studied in this thesis. A further aim was to compare different levels of aggregation. Prior to applying FA and MDS, the variation in each variable is analyzed separately. FA represents an intermediate step of aggregation, where variables with similar geographic and/or social distribution patterns are bundled together, while MDS gives the aggregate view that emerges when all variables are considered simultaneously. The relationship between the three different levels of aggregation and the extent to which the analyses complement each other is explored in this thesis.