An Acoustic Analysis of Vowel Pronunciation in Swedish Dialects
Leinonen, Therese

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2010

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
Chapter 7

Aggregate analysis

The previous chapter showed that groups can be identified among Swedish dialects based on some specific vowel features. Contrary to analyzing separate features, the dialectometric research tradition has emphasized the aggregate analysis which shows how dialects relate to each other when all available variables are considered simultaneously (Nerbonne, 2009). Common aggregating methods are cluster analysis and multidimensional scaling (MDS). Both methods use a distance matrix with the pairwise distances between all objects as input. MDS is a method for reducing complex distance data to interpretable low-dimensional representations, while cluster analysis produces partitions of the data. MDS is suitable for visualizing dialect continua, while cluster analysis detects dialect groups.

Cluster analysis is a relatively unstable method where small differences in the input data can result in substantially different outputs (Jain & Dubes, 1988; Nerbonne, Kleiweg, Heeringa, & Manni, 2008). In an analysis of Bulgarian dialects Prokić & Nerbonne (2008) found that only clusters that could be visually identified in a two-dimensional MDS plot were identified with high confidence by a number of clustering algorithms. Legendre & Legendre (1998, 482) argue that when only two or three dimensions are considered in MDS the analysis might fail to identify partitions that are distinguished in higher dimensions. However, when applying MDS to dialect data three dimensions generally explain at least around 90% of the total variance in the data (Heeringa, 2004; Prokić & Nerbonne, 2008). Hence, higher dimensions are unlikely to play any role in identifying group structure.

Tibshirani, Walter, & Hastie (2001) proposed a method called Gap statistic for estimating the number of groups in a data set. The Gap statistic can be used for estimating the number of significant clusters produced by any clustering algorithm. Lundberg (2005) used the Gap statistic to estimate the number of significant clusters when grouping the Swedish dialects based on acoustic analysis of the vowel in the word lat (/ɑː/ in Standard Swedish) and found three significant clusters.
The Gap statistic was applied to the present data set using the CLUTO\(^1\) software. The analysis showed that there are no well separated clusters in the data set. This result suggests that the Swedish dialects form a true continuum when it comes to an aggregate analysis of vowel pronunciation. The absence of clearly separable dialect groups is in agreement with previous research. The dialects in the Swedish language area are said to form a true continuum without abrupt borders (see § 2.2.1). Clustering methods could be applied to the data, but they are likely to produce unstable results, since any sharp division into subsets is not in agreement with the structure of the data. In this chapter, an aggregate analysis of vowel pronunciation in Swedish dialects is proposed by means of MDS. The MDS plots (for example, Figure 7.1) confirm the view that the Swedish language area is a genuine dialect continuum.

In § 7.1 MDS is described, and § 7.2 gives the results of a number of MDS analyses. The results of the aggregate analyses are summarized in § 7.3.

7.1 Multidimensional scaling

Multidimensional scaling (MDS) reduces complex distance data to low-dimensional representations and allows visualization of the distances in a low-dimensional space. Like principal component analysis (PCA, § 5.1.2) and factor analysis (FA, § 6.3.1) MDS is a dimensionality reduction technique. One difference between PCA and FA, on one hand, and MDS, on the other, is that PCA and FA analyze the full data matrix where every object is described by a number of variables, while MDS analyzes the distances/similarities between objects based on some chosen distance/similarity measure. In MDS the aim is to represent the objects in a small number of dimensions, while the exact preservation of original distances is less important than in PCA or FA (Legendre & Legendre, 1998, 444). In MDS priority is given to preserving the ordering of the objects instead of the exact distances between objects. Because of this, MDS allows us to investigate the relationships between dialects in fewer dimensions than FA. MDS is normally used to scale to two or three dimensions, since more dimensions are difficult to visualize simultaneously.

The results of MDS presented below show that MDS scaled to three dimensions explains more than 95% of the variance in the present data set. This can be compared to the FA in the previous chapter, where ten factors explained only 60.6% of the total variance in the data. However, FA explains 60.6% of the variance in the original data, while MDS explains 95% of the variance in the distance matrix. Converting the original data to pairwise aggregate distances as such reduces the amount of information in the data. The FA showed how Swedish dialects relate to each other when some specific features are considered. By reducing the data to a smaller number of dimensions than FA, MDS gives an aggregate analysis, that is, it allows

\(^1\)CLUTO: software for clustering high-dimensional data sets. By the Department of Computer Science and Engineering, University of Minnesota. <http://glaros.dtc.umn.edu/gkhome/views/cluto>
7.2. Dialect continuum

us to see how the dialects relate to each other when all variables are taken into account simultaneously.

In MDS, original distances between objects are approximated in a low-dimensional space by an iterative algorithm (Jain & Dubes, 1988). A number of algorithms for MDS have been proposed. In a study of the linguistic distances between varieties of Dutch Heeringa (2004) measured the fitness of three different MDS procedures by correlating the original distances with the Euclidean MDS coordinate-based distances, and found that Kruskal’s non-metric MDS gave the best results. In this chapter Kruskal’s non-metric MDS, as implemented in the RuG/L04² software, is being used.

For the MDS the distances between the varieties were calculated as the average distance of the 19 vowels in the data set. First, the distance for each vowel between two varieties was calculated as the Euclidean distance of the acoustic variables of vowel quality (Equation 6.1, p. 102), that is, two principal components (PCs), measured at nine different points within each vowel segment, starting at 25% of the total vowel duration and ending at 75%. Subsequently, the average of the vowel distances was calculated. At some locations not all of the 19 vowels were elicited and those sites had to be left out from the FA (§ 6.3). However, the average vowel distance between two objects can be calculated also for a fewer number of vowels without biasing the ordering of the data. Therefore, sites and speaker groups that were left out from the FA could be included in the MDS.

Results from MDS are often visualized in two- or three-dimensional coordinate systems. In this representation similar items are found close to each other, while dissimilar items are far apart. Nerbonne, Heeringa, & Kleiweg (1999) proposed a method for displaying the results of MDS of dialect data on maps. Each of the three basic colors in the RGB color model is used to represent one dimension of the MDS (see Appendix B, § B.1). Hence, each position in the three-dimensional space will be represented by a unique color (Figure B.1, p. 211). On the maps, the area of each variety is colored with the color corresponding to the position in the three-dimensional space. In this way, the positions in the MDS space are connected to the geographic locations, and the colors in the maps show how similar or distant dialects are to each other in a three-dimensional linguistic space.

7.2 Dialect continuum

In order to explore the Swedish dialect continuum, MDS was applied to three different divisions of the data. In § 7.2.1 the geographic variation is described by averaging over all speakers per site. In § 7.2.2 the data is split into older and younger speakers per site. Since this division of the data is the same as the one that was used in the FA in § 6.3, a comparison with the results of the FA allowed for an interpretation

of the MDS dimensions, which is reported in § 7.2.3. In § 7.2.4 a further division according to gender is made.

Group means were calculated for PC1 and PC2 respectively for each of the 19 vowels at the nine sampling points (see the lowest part of Figure 5.12 on p. 72). Subsequently the Euclidean distances between varieties were calculated based on the group means, and the resulting distances matrix was analyzed with MDS.

### 7.2.1 Geographic variation

In the first MDS analysis, average values for each site (that is, the averages of the approximately twelve speakers) were calculated for the acoustic variables before measuring the linguistic distances between sites and applying MDS. The distance matrix comprised the pairwise distances between all 98 sites.

Figure 7.1 shows the results of the MDS in a two-dimensional coordinate system where the gray-scale color of the dot represents the third dimension. One dimension explains 81.4% of the variance, two dimensions 93.6% and three dimensions 96.3%. It is clear that the first dimension already explains a very large part of the variance. Using more than three dimensions would only mean a small improvement of the variance explained. The plot does not show any clear clusters of sites, but there is

![Figure 7.1. Results of MDS to three dimensions of linguistic distances between sites. The first dimension is represented by the x-axis, the second dimension by the y-axis and the third dimension by the gray-scale color of the dot. Because of the density of sites close to the origin, a few labels have been omitted.](image-url)
only one big cloud, which supports the result of the Gap statistic that the Swedish dialects form a true continuum. The plot shows a concentration of sites close to the origin (intersection of the axes) and a more sporadic distribution in the peripheries.

The geographic distributions of the three dimensions can be viewed more easily in maps. Figure 7.2 displays the values of the sites on each of the three dimensions on maps. A one-dimensional color spectrum is used for displaying the values (Appendix B, § B.3). Magenta means high values and green means low values. In the first dimension, sites in Svealand (mainly Uppland) and on the Finnish south coast have high values, while mainly sites in Norrland have low values. Sites with high values in the second dimension are the South Swedish ones, but also the ones in Gotland and Överkalix (Norrbotten). Sites in the west of Norrland and Svealand have low values in the second dimension. The third dimension separates sites in Finland (magenta on the map in Figure 7.2, light shades in the plot in Figure 7.1) from the ones in Uppland (green on the map, dark in the plot), and the Gotlandic (magenta on the map, light in the plot) from the South Swedish ones (green on the map, dark in the plot). Närpes (Österbotten) has an extremely high value (white in the plot) in the third dimension.

The last map in Figure 7.2 shows the two first dimensions together, visualized with a two-dimensional color scheme (Appendix B, § B.2). As mentioned above, these two dimensions already explain 93.6% of the total variance in the data. The map displays the same thing as the positions of the sites in the plot in Figure 7.1 without taking the color of the dot in the plot into account. The sites in Uppland and on the Finnish south coast, with high values in the first dimension and intermediate values in the second dimension have a light yellowish color. South Swedish sites and sites on Gotland have intermediate values on the first dimension and high values in the second dimension, which leads to light blue colors on the map. The sites in Norrbotten also have quite high values in the second dimension, but very low values in the first dimension, which gives a darker blue color on the map. The third quadrant in the plot in Figure 7.1 (negative values on the two first dimensions) is dominated by sites in Norrland. These sites have dark colors on the map. The sites that are found close to the origin of the plot have grayish colors on the map. Many sites in Götaland have grayish colors. Skée (Bohuslän) is an outlier in the corner of the fourth quadrant of the plot (positive values in the first dimension and negative on the second), which gives a clear yellow color in the map. In the fourth quadrant other sites close to the Norwegian border are found as well.

Figure 7.3 shows the results of the MDS on a map using the full RGB color model, where each of the three dimensions is represented by a separate color; dimension one by the amount of red, dimension two by green, and dimension three by blue (see Appendix B, § B.1). In this map, the southernmost province, Skåne, forms a very coherent area with low values in the first and third dimensions and high values in the second dimension leading to green color. The separation of the South Swedish varieties from the ones on Gotland in the third dimension can be seen in colors close to cyan on Gotland. Uppland is also a very coherent area with an orange color. Red and purple colors are found mostly close to the Norwegian border. In Norrland
Figure 7.2. Results of MDS to three dimensions of linguistic distances between sites. All three dimensions visualized separately, green = low values, magenta = high values. The two first dimensions are also visualized together in the lower right map.
Figure 7.3. Results of MDS to three dimensions of linguistic distances between sites. Three dimensions visualized in one map with the RGB color model.
mostly dark green colors are found, but also other dark colors and blue. Götaland is quite incoherent with different colors from the center of the color spectrum. In Finland there is a clear difference between the sites on the south coast and the west coast. Saltvik (Åland) is much more similar to the sites in Uppland than to Finland-Swedish varieties.

The map where all three dimensions are represented shows that even if the distribution of dialectal features is continuous, some more coherent dialect areas can be detected.

### 7.2.2 Analysis based on age

In the following step age-related variation was analyzed in addition to the geographic variation. The distance matrix that MDS was applied to comprised the pairwise linguistic distances between 196 objects (2 age groups × 98 sites). One dimension explains 78.9% of the variance, two dimensions 92.3% and three dimensions 95.9%.

Figure 7.4 shows the results of the MDS in a two-dimensional coordinate system where the gray-scale color of the dot represents the third dimensions. The objects

![Figure 7.4](image-url)
form one big cloud, except for one outlier, which for some reason has an extremely high value in the second dimension. This outlier is the younger speakers of Löderup (Skåne). The labels of the objects do not fit into the plot, but for each object a letter indicates whether the dot concerns older or younger speakers. As can be seen, the second dimension mainly seems to separate the two age groups. The older speakers mostly have low values in the second dimension, while younger speakers have high values.

For getting a better picture of the distribution of the values in each dimension one-dimensional maps were created. Figure 7.5 shows the values in the first dimension of the older and younger speakers of each site. The extremely high value of the younger speakers of Löderup in the second dimension would mean that a large proportion of the color representing the second dimension would be required for representing this variety. In order to produce more separation between the other varieties the young speakers of Löderup were left out of the color visualizations. The maps show that in this analysis sites in Svealand and on the Finnish south coast are assigned low values in the first dimension while sites in Norrland have high values in the first dimension. This is roughly the reverse of the analysis per site in the previous section. However, in MDS the directions of the axes are arbitrary and may
Figure 7.6. The second dimension of MDS of linguistic distances between sites and age groups. Green = low values, magenta = high values.

be rotated (Legendre & Legendre, 1998, 445), which means that the first dimension in both analyses roughly represents the same thing. The maps of older and younger speakers look quite similar.

As already shown by the scatter plot in Figure 7.4, the second dimension mainly separates older and younger speakers. The maps of the second dimension in Figure 7.6 confirm the picture of the scatter plot. However, for some of the peripheral dialects (South Swedish, Gotland, Finland, Norrland) there is not such a large difference between older and younger speakers in the second dimension.

Figure 7.7 shows that high values in the third dimension are assigned to South Swedish sites, Gotlandic and all of the Finland-Swedish sites except for Åland. The older speakers in Norrbotten have high values, while young speakers in Norrbotten have low values in the third dimension. The lowest values in the third dimension are found in Jämtland. The third dimension could, hence, be interpreted as a peripherality dimension.

The maps in Figure 7.8 display all three dimensions of the MDS of older and younger speakers per site simultaneously using the whole three-dimensional RGB color spectrum. The two age groups are displayed on separate maps, but the colors of the two maps are comparable since they are based on one single MDS analysis
Figure 7.7. The third dimension of MDS of linguistic distances between sites and age groups. Green = low values, magenta = high values.

The difference between the map of the older speakers and the map of the younger speakers is striking. In the map of the older speakers a broad spectrum of colors is found, while the map of the younger speakers is dominated by green. This shows a large-scale on-going leveling of the Swedish dialects. The maps visually confirm the result of the t-test in § 6.2.1, according to which the linguistic distances between sites are significantly shorter for younger speakers than for older speakers.

By comparing the colors of older and younger speakers in Figure 7.8 conclusions can be drawn about which dialects are undergoing the biggest change when it comes to vowel pronunciation. For example, the sites in Finland have much more similar colors for older and younger speakers than many of the sites in Sweden. In order to get a more apparent view of which dialects that seem to be changing and which are stable, maps visualizing only the within-site-distances were created. In Figure 7.9 the distances between sites are disregarded and only the distances between older and
Chapter 7. Aggregate analysis

Figure 7.8. MDS to three dimensions of linguistic distances between sites and age groups. Both age groups were included in one single MDS analysis, and are represented within the same color spectrum making the colors of the two maps comparable with each other.

Younger speakers at each site are displayed. The map to the left shows the distances visualized with a continuous one-dimensional color spectrum (Appendix B, § B.3). The sites with the shortest aggregate distance between older and younger speakers are green, while sites with a large distance between older and younger speakers are magenta.

For getting an even more distinct picture, the sites were divided into three groups using K-means clustering. The map to the right in Figure 7.9 shows three groups obtained by clustering sites with the most similar distances between the two age groups. When three distinct groups are formed, all sites with only a relatively small distance between older and younger speakers are clear green, sites with a large distance are magenta, and all sites with intermediate distances between older and younger speakers are pure gray.

Dialects in the South Swedish area, on the islands Öland and Gotland, and in Finland are green, and hence have small average distances in vowel pronunciation.

---

3K-means is the most commonly used clustering algorithm for partitioning data. The user decides how many groups should be formed, and the algorithm partitions the most similar items into groups by minimizing the total error sum of squares (Legendre & Legendre, 1998, 349–355).
7.2. Dialect continuum

Figure 7.9. Maps displaying the aggregate distances between older and younger speakers at each site. In the map to the left a continuous color spectrum is used, where green indicates the shortest distance and magenta the largest distance. In the map to the right a more distinct picture is obtained by clustering the sites into three groups.

between older and younger speakers. The same holds for many sites around lake Vänern. These peripheral sites seem relatively stable when it comes to vowel pronunciation. Many of the more central sites around and south-west from Stockholm are gray or magenta, which suggests a big ongoing change in vowel pronunciation. This is also the case for sites close to Göteborg on the west coast. In Norrland there are sites of all three types: some dialects show a large ongoing change, some an intermediate change, and some are relatively stable.

In the map of the younger speakers in Figure 7.8 it looks as if there is almost no variation in vowel pronunciation between younger speakers. This is not entirely true. The variation between younger speakers is only so much smaller than between older speakers and between the two generations that only a small part of the color spectrum can be used for showing the differences between younger speakers at different sites.

In order to be able to visualize dialectal differences within the younger age group, MDS was also applied separately to the older and the younger speakers. That is, two separate distance matrices were analyzed, one with the distances between older speakers at all sites and one with the distances between younger speakers at all sites.
Chapter 7. Aggregate analysis

Figure 7.10. MDS to three dimensions of linguistic distances between sites for each age group separately. The maps are based on two separate MDS analyses, so that the full color spectrum is used in each of the maps. The colors are not comparable across the two maps.

The analysis of only older speakers included 98 sites and the amount of explained variance was 82.1% for the first dimension, 93.4% for two dimensions and 95.7% for three dimensions. The analysis of only younger speakers included 97 sites (younger speakers in Löderup were left out) and the amount of explained variance was 83.8% for the first dimension, 92.4% for two dimensions and 96.1% for three dimensions.

Figure 7.10 shows the maps with results of MDS applied separately to the older and younger speakers. Since two separate analyses are displayed in the two maps, the colors of the maps have to be interpreted independently. The maps are similar to the ones in Figure 7.8, but the colors in each map are more distinct. When the whole color spectrum is used for each age group separately it becomes clear that there are differences across sites among the younger speakers, which could not be distinguished in Figure 7.8. Moreover, the geographic pattern is quite similar for older and younger speakers. So even if the dialectal differences in vowel pronunciation are larger in the

For example, Gotland has completely different color in the two maps. This does not mean that there is a large difference in vowel pronunciation between older and younger speakers on Gotland, it is only an effect of the full color spectrum being used to visualize the distances within each age group.
older generation than in the younger, the geographic distribution of dialectal features remains more or less the same.

Some differences in the geographic distributions can also be found. For example, the dialects in Norrland are more coherent in the younger age group than in the older. Figure 7.9 showed that some dialects in Norrland are relatively stable, while others have a large linguistic distance between older and younger speakers. In Norrland the most divergent dialects seem to be changing the most, and thereby a more uniform spoken variety of Norrland is emerging. This can be seen as regionalization of the dialects, since the most divergent dialectal features seem to be disappearing while some other features that distinguish Norrland varieties from other Swedish varieties are preserved.

7.2.3 Interpreting MDS dimensions

Because MDS is based on a distance matrix with pairwise aggregate distances between varieties, MDS does not offer any direct way to interpret which of the original linguistic variables that have caused the distribution in the extracted dimensions. In factor analysis (FA), however, the loadings indicate the correlation between original variables and extracted factors. In § 6.3 the average acoustic values per age group per site were analyzed by means of FA. The same division of the data into older and younger speakers per site was used in the MDS in § 7.2.2, which makes a comparison between MDS and FA possible.

The values of the objects in the three dimensions of the MDS were correlated with the scores of the FA using Pearson’s correlation. The correlations are displayed in Table 7.1. The first dimension corresponds well to the first factor ($r = 0.921$) and the second dimension correlates highly with the third factor ($r = 0.840$). The third dimension does not show very high correlations with one single factor, but seems to be a combination of several of the factors. Most strongly the third dimension correlates with the fifth ($r = 0.468$) and the ninth ($r = -0.502$) factor.

Since the FA includes loadings that tells us which of the linguistic variables are connected to each factor, a linguistic interpretation of the MDS dimensions can be inferred from the correlations with the factors. Based on the loadings of the FA, the

| Table 7.1. Pearson’s correlations between factors and MDS dimensions. Insignificant correlations ($p > 0.05$) are indicated by a dash. |
|-----------------|-----|-----|-----|
| factor1        | 0.921 | -0.227 | - |
| factor2        | 0.213 | 0.302 | 0.313 |
| factor3        | 0.224 | 0.840 | - |
| factor4        | -    | -    | 0.195 |
| factor5        | -    | -    | 0.468 |
| factor6        | -    | -    | - |
| factor7        | -    | -0.160 | 0.176 |
| factor8        | -    | 0.177 | -0.320 |
| factor9        | -    | -0.162 | -0.502 |
| factor10       | -    | -    | -0.218 |
conclusion can be drawn that the first dimension of the MDS is largely based on differences in the PC2 values of front vowels (compare with the loadings of the first factor of the FA in Table 6.5, p. 111), which is assumed to be connected to voice quality differences.

The second dimension of the MDS correlates highly with the third factor of the FA, and hence has to do with differences in the PC1 values of front mid vowels (Table 6.7, p. 116). The FA showed that these vowels are lowered by younger speakers in many areas.

The third dimension is more unspecific and shows the effect of several variables, which have different geographic distribution patterns according to the FA. It seems that varieties with extreme scores in the third dimension of the MDS are not necessarily linguistically very similar to each other, but they are characterized by dialectal features that make them divergent from more central varieties.

The effect of the second factor of the FA, which distinguishes South Swedish varieties from the rest, is spread over all three dimensions of the MDS (significant correlations between 0.2 and 0.3 with all three dimensions).

The sixth factor of the FA, which detected only very subtle differences between dialects (see § 6.3.7), does not correlate significantly with any of the MDS dimensions.

7.2.4 Analysis according to age and gender

In the final MDS analysis the data from every site were divided into four groups: older men, older women, younger men, younger women. This resulted in a 390 × 390 distance matrix (speaker groups with less than 15 vowels were not included, see Appendix A). One dimension explains 82.5% of the variance, two dimensions 93.0% and three dimensions 95.9%.

Figures 7.11–7.13 display one-dimensional maps of the three first dimensions separately, with one map for each speaker group. Green color indicates low values and magenta high values (see Appendix B, § B.3).

The solution is similar to the one in § 7.2.2, only with the second and third dimensions reversed. Low values in the first dimension (Figure 7.11) were assigned mainly to Svealand and the Finnish south coast, even though the geographic distribution is not completely coherent.

The second dimension (Figure 7.12) mainly separates older and younger speakers. At most sites, the younger speakers have low values, while high values are found among older speakers (the inverse of the second dimension of the MDS in § 7.2.2). In the second dimension the older women have somewhat higher values on average than the older men.

As in the analysis in § 7.2.2, the third dimension (Figure 7.13) could be called a peripherality dimension. South Swedish, Gotlandic and Finland-Swedish sites and the older speakers in Norrbotten have low values, while speakers in Jämtland have high values (the inverse of the third dimension of the MDS in § 7.2.2).

Figure 7.14 displays the four maps that combine all three dimensions using the RGB color model. All three colors were reversed to obtain a color representation
7.3. Conclusions of the aggregate analysis

In this chapter the relationships between varieties of Swedish were analyzed using aggregate linguistic distances based on acoustic measurements of 19 vowels. The Gap statistic showed that the data cannot be partitioned into groups, but the Swedish dialects form a continuum without abrupt borders. This is in line with previous descriptions of the Swedish dialects. For visualizing the continuum-like relationships between varieties of Swedish, multidimensional scaling (MDS) was used. Five different analyses were carried out using MDS. In § 7.2.1 the linguistic distances between sites were analyzed. In § 7.2.2 the data were divided into older and younger speakers per site and three analyses were made: one including both older and younger speakers, one with only older speakers, and one with only younger speakers. In § 7.2.4 a further division according to gender was made and an analysis which included four speaker groups per site was carried out.

The analysis of linguistic distances between sites showed that even if the distribution of dialectal features is continuous, some more coherent dialect areas can be detected.

The analysis of the two age groups in the data set showed that the dialectal differences are considerably smaller in the younger generation than among older speakers. The effect of dialect leveling in apparent time is large. However, the geographic distribution of dialectal features is not changing much, so that the main dialect areas remain the same. This can be interpreted as regionalization of the dialects, since it seems that dialectal features that are characterizing the larger dialect regions are still being preserved at the same time as the overall linguistic distances become smaller.

An analysis of the distances between the two age groups within each site showed that the central dialects, close to the biggest cities, seem to be changing the most, while many of the more peripheral dialects are relatively stable when it comes to an aggregate analysis of vowel pronunciation. This can seem surprising, because similar to the ones in the previous sections (Figure 7.8 and Figure 7.3). From these maps, it is obvious that the differences in vowel pronunciation are larger between the two generations than between men and women. The map of the older women is darker than the one of the older men, mainly due to the second dimension. As in the analysis in § 7.2.2, we can see that the dialectal differences are smaller in the younger generation than in the older, disregarding the variation across genders. The maps of both the younger men and the younger women are dominated by green, while the maps of the older speakers show a broader color spectrum. The maps of the older speakers show roughly the same geographic patterns that were detected in the two previous sections, while the color spectrum used for the younger speakers is so narrow that almost no geographic patterns can be detected. For the site Skee (Bohuslän) it can be noted that the younger men have a red color, like older speakers, while the young women have a color much more similar to the surrounding dialects.
in a dialect leveling situation you might expect that the most divergent peripheral dialects would be converging to more central and standard-like varieties. In order to understand why the central dialects are changing the most this result of the aggregate analysis has to be studied in relation to the change in each of the variables. This is discussed further in § 8.2.2.

A further division according to gender showed that the differences between older and younger speakers is much larger than the difference between men and women. In the older generation there seems to be more gender-related differences than in the younger generation.

The results obtained by MDS could be correlated with results from the factor analysis (FA) in the previous chapter. This showed that the first dimension, which separates Svealand and south Finland-Swedish varieties in all the MDS analyses, is largely explained by the PC2 values of front vowels. The second dimension, the one producing the largest distance between older and younger speakers, is mainly an effect of the lowering of front mid vowels by younger speakers. The third dimension of the MDS is a peripherality dimension separating peripheral areas in the Swedish language area, like Skåne, Gotland, Finland, Norrbotten and Jämtland. Variables characterizing peripheral dialects were spread on several factors with different geographical distributions in the FA.
Figure 7.11. The first dimension of MDS of linguistic distances between sites and speaker groups based on age and gender. Green = low values, magenta = high values.
Chapter 7. Aggregate analysis

Figure 7.12. The second dimension of MDS of linguistic distances between sites and speaker groups based on age and gender. Green = low values, magenta = high values.
7.3. Conclusions of the aggregate analysis

Figure 7.13. The third dimension of MDS of linguistic distances between sites and speaker groups based on age and gender. Green = low values, magenta = high values.
Figure 7.14. Results of MDS to three dimensions of linguistic distances between sites and speaker groups based on age and gender, visualized with the RGB color model. All speaker groups are represented within the same color spectrum making the colors of the maps comparable with each other.