Rhythm in Stockholm’s two working-class varieties: Separate models predict intervocalic durational contrast

Nathan J. Young
Queen Mary, University of London
nathan.j.young@qmul.ac.uk

Abstract
This study shows that two distinct social models predict speech rhythm variation – measured by the normalized pairwise variability index of vowels (nPVI-V) – for Stockholm’s two working classes. The non-white working-class variety (multiethnolect) has less intervocalic durational contrast than the speech of elites (41–51 vs. 49–57) and correlates with the speaker’s neighborhood diversity. Incremental increases in neighborhood diversity correlate with incremental decreases in nPVI-V. The white working-class variety has more intervocalic durational contrast than the speech of elites (53–61 vs. 49–57) and correlates to occupational status. Incremental decreases in occupational status correlate with incremental increases in nPVI-V.

The data comes from 31 male Stockholmers, ages 24–49, who read aloud a passage with 285 vocalic elements. Thirteen self-identify as white ‘Swedes’: five working class, eight upper-middle class (‘elites’). Eighteen self-identify as non-white ‘immigrants’: five working class, seven lower-middle class, six upper-middle class (‘elites’). Twenty-eight were born in Sweden; three arrived before age four. They hail from five neighborhood types that are representative of Stockholm’s geographic ethnic distributions.

The findings add Swedish multiethnolect to a growing list of contact varieties with less intervocalic durational contrast than their heritage counterparts. The findings also nudge our field’s discussion of rhythm away from second-language acquisition to the social domain of race and class. At the same time, a new research question emerges whether intervocalic durational contrast is a sociolinguistic variable in its own right or a byproduct of segment-level variation.

Index Terms: multiethnolect, nPVI, Rinkeby Swedish, Stockholm Swedish, variationist sociolinguistics, vernacular

1. Background

1.1. Ethnosocial dialects of Europe and Stockholm

Much of Europe is witnessing the development of new sociolinguistic standard and referred to as contemporary urban vernaculars [1]. Stockholm’s multiethnolect is part of this linguistic phenomenon that has been exclusively seen (thus far) in cities of European countries that ran postwar guest-worker programs. These include Berlin [2], Copenhagen [3], Gothenburg [4], Hamburg [5], London (Multicultural London English, ‘MLE’) [6], Malmö [7], Mannheim [8], Oslo [9], Paris [10], Rotterdam [11], and the South Midlands [12].

1.2. Adult speakers accentuate class and age strata

Most of the aforementioned works are investigations of youth language. However, a growing number of studies have examined sociolinguistic variation among British-Asian adults, young and old [1, 13], many of whom are MLE speakers. Research on Germany’s Turkish diaspora has also improved our understanding of adult ethnosocial linguistic practice [8].

Such research is important because the repertoires typically ascribed to youth have already matured into ethnosocial dialects that are spoken by a diverse non-white class of working-age adults [1, 14, 15, 16]. Few researchers have investigated this ‘adult question’ in Sweden. An ethnographic study of Assyrian-Swedish men in Stockholm (n=3) found that they style-shifted from ‘native’ speech at work to ‘non-native’ speech in private [17], echoing [1] and [13]. The findings challenge outdated notions in Sweden about L2 deficiency and imply that the ‘non-native’ register was actually Stockholm’s new ethnosocial dialect. A later study found that six speakers from multiethnic suburbs were assessed differently by listeners in a perception experiment: two were said to be ‘neutral’ and ‘Swedish’ while four were assessed as ‘rough’ and ‘non-Swedish’ [16]. This was attributed to the former two style-shifting like the Assyrian-Swedes of the first study [17].

1.3. Geography matters in a segregated Stockholm

In this paper, multiethnolect [18] refers to the variety often spoken in Stockholm’s multiethnic neighborhoods. Sundry terms circulate, but Rinkeby Swedish [19, 20, 21, 22] remains the most well known, although it no longer is solely bound to the immigrant suburb, Rinkeby. It has other 21st-century geographic indexicalities, including the outer northwestern and southwestern suburbs [21, 23] (Figure 1).

Like in other segregated European cities [24], ‘neighborhood matters’ in Stockholm and is stereotyped to certain social classes and types. In this manner, multiethnolect is like other varieties: working-class Södermalm (aka Ekensnack) is tied to Södermalm and the inner southern suburbs; ‘posh’ Stockholmian is tied to Danderyd and Östermalm [15, 23, 25] (Figure 1).

1.4. Two varieties in symbolic opposition

Södersnack was spoken by working-class Stockholmers for most of the twentieth century until other speech styles began to emerge, including multiethnolect. Although both Södersnack and multiethnolect ‘belong’ to Stockholm’s working class, the similarities end there. The former is often attributed to white speakers in the inner southern suburbs and the latter to non-white speakers in the outer northwest (Sw. Västerort; En. Westside) and southwest (Sw. Söderort; En. Southside). The former is often described by the media with endearment [26] and the latter with alarm [20, 22].

Manufacturing’s decline and migration’s rise have stoked tension between the white working class and the multiethnic working class, aiding the rise of the national socialist Sweden Democrats (Sw. Sverigedemokraterna) [27, 28]. Indeed, one striking finding of the ethnographic part of this study was
that all white working-class participants vote Sweden Democrat whereas the multiethnolectal participants vote Left (Sw. Vänsterpartiet) or Social Democrat (Sw. Socialdemokraterna).

Two studies have examined multiethnolect features in juxtaposition to white working-class speech [6, 10], but to date, no study has examined systematic variation in Stockholm. As far as prosody is concerned, one pilot study has examined multiethnolect (see 2.2) [16]. Prosody in Södersnack has never been studied, but some of its long vowels are more diphthongal than the standard [25, 29], which might mean more intervocalic durational contrast than in multiethnolect and standard styles.

2. Speech rhythm

2.1. Defining the variable

The study of prosodic rhythm emerged as a way to measure degrees of isochrony, a now-disproven concept [30] (see [31] for a more complete review). It was assumed that there is “a straightforward relationship between duration and abstract phonological categories, such as syllable structure, vowel weight and vowel reduction patterns” [31]. Therefore, languages with low intervocalic durational contrast were assumed to be “syllable-timed” and languages with high intervocalic durational contrast were assumed to be “stress-timed”.

This study intentionally disengages from the taxonomy of timing per se. Its catalyst are the many works that describe Rinkeby Swedish as ‘staccato’ or ‘jerky’ (Sw. stötig) [15, 16, 19, 22]. Plainly stated, this study seeks to investigate whether “the regular alternation of strong and weak elements [32]” is in itself a sociolinguistic variable in late-modern Stockholm.

One could speculate that ‘strong’ and ‘weak’ would manifest themselves in F0 and intensity. However, these two features rely heavily on the progression of time in order to operationalize contrast. This claim is supported by research that has found that speech prominence – while influenced by F0 and intensity – is nonetheless highly reliant on duration [33, 34]. Therefore, this paper defines rhythm as local durational alteration.

Aggregate or global metrics [35] like percentage of vowel duration to segmental duration [36], standard deviation [36], and variation coefficients [37, 38] do not model the local essence of strong and weak elements [35]. Therefore, I use a local metric [35] known as the normalized pairwise variability index for vowels (nPVI-V) [39]. It creates a numerical representation for the durational alternation of consecutive vowels and normalizes for speech rate:

\[ nPVI = \frac{|d_{n+1} - d_{n}|}{d_{n+1} + d_{n}} \cdot 100 \]  

where \( d_{n+1} \) = duration of the first vowel in the pair
where \( d_{n+1} \) = duration of the second vowel in the pair

2.2. Rhythm in contact varieties and Swedish

Most analyses of rhythm in contact varieties have conceptualized contrast as durational, and many have used the nPVI-V algorithm to model this (cf. [3, 7, 10]). Most of these studies show reduced contrast for contact varieties than their heritage counterparts [6, 35, 39, 40, 41, 42] (cf. [35, 41]).

Swedish prosody has been examined with nPVI-V twice. In a small pilot study of male speakers (\( n = 8, 50 \) vocalic elements), speech assessed as multiethnolectal had lower intervocalic durational contrast (\( n = 4: 39.2, 40.6, 45.3 \& 47.3, \) respectively) than speech assessed as the received standard (\( n = 4: 52.6, 52.7, 54.8 \& 54.9, \) respectively) [16].

Another study of female speakers (\( n = 10, 120 \) vocalic elements) showed that standard Central Swedish has an nPVI-V that ranges between 46.0 and 57.4 and that L2 Swedish spoken by L1 Estonians ranges between 44.7 and 53.0 [42]. These studies were preceded by one other that measured the duration of vowelophonemes, hypothesizing that long vowels shorten and short vowels lengthen in Malmö multiethnolect. Both hypotheses were invalidated [7].

3. Definitions and research questions

DEFINITIONS: In this paper, the upper-middle class are sometimes referred to as elites. I use white instead of ‘Swede’ or ‘native’ and non-white and of color instead of ‘immigrant’ or ‘non-native’. This very topic has been tangentially [11, 14, 16, 17, 20] and explicitly [43, 44] explored elsewhere but will not be handled here.

RESEARCH QUESTION A: Is the intervocalic durational contrast of Stockholm’s non-white working-class variety, i.e., multiethnolect, lower than both the elite variety and the white working-class variety?

RESEARCH QUESTION B: Is the intervocalic durational contrast in Stockholm’s white working-class variety higher than both the elite variety and multiethnolect?

4. Method

4.1. Speakers

Thirty-one adult male speakers, ages 24–49, participated in the study, recruited via the snowball method. For the purposes of controlling for social class and ethnicity, women were excluded (see discussion in 8.4). The speakers are categorized by social class according to the Swedish Standard Classification of Occupations 2012 (SSYK) [45].

Thirteen self-identify as white ‘Swedes’ (Sw. svensk); five working class, eight upper-middle class. Eighteen self-identify as non-white ‘immigrants’ (Sw. invandrare: five working class, seven lower-middle class, and six upper-middle class. Eleven different ethnic heritages are represented: Azeri (1), Chilean (1), Egyptian Arab (1), Iranian (3), Kurdish (3), Mimbriel (1), Peruvian (1), Qashqai (1), Somali (1), Tigrinyan (4), and Turkish (1).

All speakers were born in Sweden except for three who arrived before the age of four. All speakers reported Swedish as their strongest language for speaking and understanding.

4.2. Speech material and phonetic analysis

The participants read an adapted version of The Circus (Sw. Cirkusen) [46], a reading passage that includes multiple examples of all Swedish phonemes and pitch accents. For the analysis, I removed disfluencies and hesitation markers, resulting in data ranging from 215 to 277 vocalic elements, depending on elisions. Earlier benchmarks for the number of vocalic elements analyzed are \( n = 237 \) [6], 80 [39], and 200 [40].

The final foot before a pause was included in the calculation [6, 35, 38, 39], cf. [40, 41]. Syllable-coda /t/ and /j/ were included as part of the vowel, whereas intersyllabic /t/ and /j/ were not [40]. Breath groupings were delineated by pauses of 70 ms or more [40] (cf. 150 ms [35]).

\(^{1}\) Some elites reported English as equally strong.
4.3. Social predictor variables

Age, social class and race/ethnicity are typical social factors to test. In Stockholm’s context, neighborhood is also important, given the strong geographic indexicality of the city’s varieties.

Birth year was used for age. For social class, the status rank of each speaker’s occupation was used [47]. Status rank captures the incremental differences in symbolic capital between, say, a waiter and a demolition worker (both working class) or a doctor and a portfolio manager (both upper-middle class).

Neighborhood was broken into five classes that correspond to five tiers of racial diversity: westside working-class communities of color (e.g., Rinkeby, Tensta: 90–100% residents of migrant origin), southside working-class communities of color (e.g., Botkyrka, Skärholmen: 80–90% migrant), mixed working-class communities (e.g., Fisksättra, Hässelby, Rågsved: 40–60% migrant), white working-class communities (e.g., Bagarmossen, Farsta: 20% migrant), Central Stockholm and affluent villa suburbs (e.g., Danderyd, Saltsjöbaden: 0–10% migrant). The average percentage of residents of migrant origin for each neighborhood was entered into the analysis: 95, 85, 50, 20, and 5 for the above, respectively.

For race, white and non-white were entered as binary dummy variables 0, 1 (see definitions in Section 3).

All four predictors were tested for orthogonality with a Pearson product moment correlation (n = 31). Occupational status ~ race are orthogonal (R = −0.22), but each of them correlated with neighborhood diversity (R = 0.69 and R = −0.68, respectively), which implies that neighborhood is a robust factor for capturing the intersection of race and class. Age ~ neighborhood (R = 0.55) and age ~ race (R = −0.57) have medium correlation, which means they can be entered into a multiple regression together but interpreted cautiously. Occupational status ~ age have low correlation (R = 0.16).

5. Overview of Results

Figure 1 presents all the data with occupational status on the x axis and nPVI-V on the y axis. Race and neighborhood are superimposed to provide visual insight into their collinearity with neighborhood. Two models emerge: one for the relationship between elite speech and non-white working-class speech, and the other for the relationship between elite speech and non-white working-class speech. The positive-sloped least-squares regression line represents the nPVI-V trend between elites and the multiethnic working class; the negative-sloped one represents the nPVI-V trend between elites and the white working class.

Figure 1 shows that speakers from neighborhoods with lower percentages of immigrant residents tend to speak with more durational contrast, albeit with one exception. The white working class almost entirely inhabits neighborhoods with ca. 20% residents of migrant origin. They are marked with * in the plot. Despite their moderate proximity to residents of migrant origin, they appear to defy the general trend for nPVI-V. For that model, occupational status seems to be a more relevant predictor than neighborhood diversity.

6. Question A - Multiethnolect

After removing the five white working-class speakers (denoted with *) and running the following linear regression, nPVI-V ~ Neighborhood + Birth_year (n = 26), neighborhood diversity shows a significant effect on intervocalic durational contrast (p < 0.01). Table 1 displays the results. The model indi-
cates that a 1% increase in a speaker’s neighborhood diversity results in a 0.070 drop in nPVI-V. In other words, being from a multiethnic neighborhood correlates with lower intervocalic durational contrast. Age was not a significant predictor.

**Research Question A** appears to be confirmed. The intervocalic durational contrast of Stockholm’s non-white working class speech (multiethnolect) – defined by non-white speakers from immigrant-dense neighborhoods – is lower than that of non-white speakers from mixed neighborhoods and white and non-white speakers from affluent white neighborhoods.

7. **Question B - White working-class speech**

A clear trend emerges in Figure 1 if one excludes the speakers from neighborhoods with 50, 85, and 95 diversity percentages. With only two neighborhood diversity tiers remaining – 20% and 5% – occupational status emerges visually as a more relevant predictor. The trend suggests that low occupational status coincides with higher nPVI-V’s, which was tested with a new regression model whereby all non-white working-class participants were removed: $nPVI-V \sim \text{Occup}\_\text{status}\_\text{rank} + \text{Birth}_\text{year}$ ($n = 17$).

Table 2: Regression analysis – Occupational status and age as predictors for nPVI-V: non-white working class excluded.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>t-value</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Occup_status_rank</td>
<td>0.072**</td>
<td>0.018</td>
<td>4.06</td>
<td>$p &lt; 0.001$</td>
</tr>
<tr>
<td>Birth_year</td>
<td>-0.160**</td>
<td>0.065</td>
<td>-2.56</td>
<td>$p &lt; 0.05$</td>
</tr>
<tr>
<td>Constant</td>
<td>368.081**</td>
<td>129.004</td>
<td>2.30</td>
<td>$p &lt; 0.50$</td>
</tr>
<tr>
<td>R²</td>
<td>0.707</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variance inflation factor</td>
<td>3.413</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table 2 shows that occupational status has a significant positive effect on nPVI-V when non-white working-class speakers are removed from the data set. Birth year emerges as significant and has a mild negative effect. The model indicates that a singular drop in occupational status rank will result in a 0.072 higher nPVI-V. For every decade younger, the nPVI-V of a speaker would decrease by 1.6.

**Research Question B** appears to be confirmed. The intervocalic durational contrast in Stockholm’s white working-class speech is higher than that of white and multiethnolect speech.

8. **Discussion**

8.1. **Descriptive findings**

The findings suggest that for male speakers, multiethnolect ranges between 41 and 50 for nPVI-V, and elite Swedish appears to range between 50 and 57, which is consistent with earlier findings [16, 42]. White working-class Swedish ranges between 53 and 60.

8.2. **Addressing the research questions**

Two separate models operate. One model is the relationship between the speech of elites and that of the multiethnolect working class, in which geography appears to be especially relevant. An increase in neighborhood diversity results in lower intervocalic durational contrast than of those who have high-status work and live in affluent neighborhoods. This is particularly the case for westside and southside communities. Non-white working-class speakers from mixed communities appear to have an nPVI-V that is consistent with speakers from affluent communities.

The other model is the relationship between the speech of elites and white workers, in which social class appears to be more relevant. A decrease in occupational status results in higher intervocalic durational contrast than of those who are high-status and live in affluent neighborhoods. For the white workers, the increased neighborhood diversity does not appear to reduce intervocalic durational contrast.

8.3. **White working-class results: legacy or innovations?**

The results for the white working class are difficult to interpret because the subsample is small ($n = 5$). Are they speaking a version of Stockholm’s older Södersnack or are they engaging in an oppositional speech act in reaction to their multiethnic rivals? More speakers should be added to the corpus to inform the analysis.

8.4. **Future directions**

Examining other linguistic variables that co-occur with speech rhythm can help paint a more accurate picture of what actually constitutes these repertoires. Examining other speech styles aside from reading, such as casual speech, can also illuminate to what degree these findings constitute style-shifting and to what degree they actually reflect late-modern vernacular style in Stockholm. Furthermore, the use of a single occupational status metric [47] may not be the best way to capture social class. Metrics that take into account parental occupation, education, occupational status, income and taste should also be considered.

It is unclear whether intervocalic durational contrast is a variable in its own right or whether these results are a byproduct of diphthongs in white working-class speech and monophthongs in multiethnolect. The findings call for a phonological investigation into the role of underlying segmental variation. A final point is that the speech of women in Stockholm is not included in this study. Until data on women’s speech is collected and analyzed, no full conclusion on prosody in late-modern Stockholm Swedish can be reached.

9. **Acknowledgements**

This work was supported by Queen Mary, University of London, the British Economic & Social Research Council, the Centre for Research on Bilingualism at Stockholm University, Sven och Dagmar Saléns stiftelsen, and the guidance of Devyani Sharma, Erez Levon, and Erik Thomas. All errors are my own.

10. **References**
