ETAP4 Special Issue

Language
and Speech

Language *and* Speech
1–37
© The Author(s) 2021
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0023830920969727
journals.sagepub.com/home/las
SAGE

The Sociolectal and Stylistic Variability of Rhythm in Stockholm

Nathan J. Young
Centre for Research on Bilingualism, Stockholm University, Sweden

Abstract

The question of "staccato" rhythm in Stockholm's multiethnolect is investigated by comparing nPVIV measurements of the speech of 36 adult male speakers. The men, ages 24–43, come from a stratified sample of social classes and racial groups. Three contextual styles were recorded and analyzed: informal, formal, and very formal. The distribution of nPVIV values in informal speech across class and racial group indicates that speech rhythm splits three ways: low-alternation "staccato" rhythm among the racialized lower-class men, high-alternation rhythm among the white lower-class men, and an intermediate level of rhythm among higher-class men, regardless of racialized category. The "staccato" low-alternation feature is also less stylistically sensitive than the high-alternation feature, implying that the latter is a more established feature than the former. Further, the "staccato" feature is more stylistically sensitive among younger speakers than older speakers, implying an ongoing change from *indicator* to *marker* status. For all speakers, age has a stable main effect, which means that younger speakers, independent of racial group and class, have lower alternation than older speakers. Implied here is that low-alternation is a change from below that originates within the racialized working class. While it may be incrementally transmitting into the wider speech community, the white working class is the most resistant to its incursion.

1 Introduction

This study sets out to investigate the "staccato" variant in the prosody of male speakers of Stockholm's multiethnolect, and in doing so, the results extend beyond this single variant. While finding evidence in support of said staccato feature, the findings also show that the social stratification of rhythm is horizontal rather than top-to-bottom. This is to say that white working-class men have the highest rhythmic alternation in their prosody, and racialized[1] working-class men have the lowest alternation in their prosody. Speech rhythm among the upper and middle classes fall between the two.

These findings join in on two separate discursive threads that have been simultaneously running for some time. One thread is grounded in local descriptive work, and the other in global linguistic theory. Locally, and ever since Kotsinas' (1988a) first survey of Rinkeby Swedish, scholars and

Corresponding author:
Nathan J. Young, Centre for Research on Bilingualism, Institute for Research on Bilingualism, Stockholm University, Stockholm, 106 91, Sweden.
Email: nathan.young@biling.su.se

laymen alike have been commenting on the staccato impression of the new youth vernacular in Stockholm (as well as in Malmö and Gothenburg). In addition to these accumulating impressionistic accounts, two small investigations of rhythm have been conducted (Bodén, 2007; Young, 2018b), but neither have been able to fully substantiate the staccato claim.

As it pertains to linguistic theory, since the early 1980s, a global literature on speech rhythm has been growing and maturing. This literature has taken aim at rhythmic typologies for the world's major languages as well as the contact varieties associated with many of those languages. What has not been seen, however, is an investigation of rhythm as a *sociolinguistic variable* in the Labovian sense of the word. Specifically, the rhythmic variants in contact scenarios have not been examined as changes from below whose approximate salience and age can be assessed by means of stylistic sensitivity. This theoretical paradigm is a main point of departure for the present investigation.

The analysis will show that rhythmic alternation of the white working class is more stylistically sensitive than of the racialized working class, which aligns with the interpretation that the former is a legacy feature of Stockholm's working-class Södersnack and the latter a feature of the newer multiethnolect. For all social groups investigated, low-alternation rhythm appears to be a feature of younger men, and high-alternation rhythm a feature of older men, which is to imply a systematic change in apparent time toward lower-alternation prosody all around. This main effect for age, however, is complicated by the aforementioned horizontal stratification of the race and class nexus.

The research questions addressed by this study are (a) whether the speech of Stockholm's racialized working class has lower rhythmic alternation (staccato) than the speech of other male groups in the city, (b) to what degree the speech-rhythm variants are stylistically sensitive, and (c) what conclusions can be made about the trajectory of rhythm in the city as a whole, including in other varieties besides the urban vernacular.

2 Background

It became apparent during the ethnographic portion of this study that three distinct social groups operate in Stockholm today—(a) a new racialized working class, (b) the established white "Swedish" working class, and (c) a diverse middle and upper-middle class whose cosmopolitan aspirations somewhat serve to neutralize racial and ethnic boundaries. This study investigates speech rhythm among male speakers within each of these three groups. The first group generally speaks varying "intensities" of Stockholm's multiethnolect; the second group generally uses linguistic features from Stockholm's industrial-era Södersnack; the third group speaks standard Central Swedish with varying degrees of upper-class feature bricolage.[2] This background review will focus on the two working-class groups, their history, and their speech forms–since they are the predominant locus of the stratification of rhythm in Stockholm.

2.1 Multiethnolects, racialization, and late modernity

As I propose above, Stockholm has a working class that has divided itself along racialized lines. A key social category used in this analysis is whether participants self-identify as *svensk* or *invandrare*, two terms that literally translate as "Swedish" and "immigrant," respectively. Although the terms ostensibly refer to properties of national and migrant origin, they are in fact used in Sweden (in the 2010s) as terms of race and racialization and translate into "white" and "non-white," respectively. When one wishes to colloquially refer to actual immigrants or speakers of late-acquired L2 Swedish, the term *import* is used (no speaker in this study would qualify as an *import*). Throughout this paper, *svensk* and *invandrare* are used emically to refer to their colloquial racial connotations instead of actual national or migration background.

In order to understand how these two identities figure into the analysis of rhythm, it is necessary to provide a review that builds a sociological argument for their importance. Unlike, for example, social class, the concept of "race" in the European context is often questioned. Furthermore, the term "immigrant" can lead readers to forget that these speakers are early acquirers of Swedish, with all of them reporting Swedish to be their strongest language and most of them having an age of onset between 1 and 2 years (see Section 3.1).

The linguistic development in Stockholm, as in many cities that have witnessed significant post-War migration, is a uniquely late-modern phenomenon. *Late modernity*, itself a term often left underdefined, is characterized by Wacquant (2008) as the current epoch in which both manufacturing and the welfare state have starkly weakened. The Poststructuralist school has referred to the era as "post-Fordist," and Bauman (2000) has famously referred to it as "liquid modernity." One key way in which late modernity is tightly connected to the present linguistic development is via the liberalization of schools (Rampton, 2006). Multiethnolects did not just arise from the banal confluence of contact scenarios; rather, the establishment of independent charter-school markets in numerous European countries like the UK and Sweden has led to the hyper-concentration of minority pupils in some schools and majority pupils in others (see Forsberg, 2018, for an excellent analysis). Growing income stratification and residential segregation have worked in tandem with decades-long school segregation to racialize the social-class hierarchy, something that is now said to be a signature feature of the *European* strain of late-modernity (Hesse, 2007; Lee, 2010; Lentin, 2008; Lentin & Titley, 2011).

In Stockholm, this demographic subgroup has developed its own linguistic variety after more than 40 years of school segregation, social exclusion, and relegated suburban enclosure. Critical Race theorists like Hübinette, Hörnfeldt, Farahani, and Rosales (2012) have therefore also proposed that "*invandrare*" is in fact a racialized euphemism for what Mulinari and Neergaard (2004) have referred to as Sweden's racialized working class. The term *Swedish multiethnolect*, while used in this paper, is actually inadequate because it neither indexes the variety's raced nor classed attributes (Rampton, 2011). Rather than being the lect of multiple ethnicities, the lect is the variety of an underclass for whom ethnic differences have been *erased* amid the engrossing salience of their phenotypic markedness within the majority population.

The above discussion is not just important for understanding the current racialized system that has emerged in Stockholm. It is also important for framing the research of multiethnolects in a more general sense. Scholars too often assume there to be L1 transfer on L2-Swedish while forgetting two important points. First, the speakers of these varieties are early acquirers of the majority language, well within Lenneberg's (1967) Critical Period, and many have the majority language as their home language. Second, working-class youth—monolingual or not—have always been the main innovators and drivers of language change (Labov, 2001). Sometimes these innovations are appropriated from a more endogenous feature pool like t-glottaling in Norwich (Trudgill, 1988) or PRICE-centering in Martha's Vineyard (Labov, 1963), and other times they are appropriated from a more exogenous feature pool like KIT-raising in Los Angeles (Mendoza-Denton, 2008, pp. 230–264). Despite these two points, the mechanical effects of bilingualism are sometimes seen as the default explanation for variation when L2-speakers enter the scene, while social factors are sidelined (Bodén, 2007; Kotsinas, 1988a).

2.2 Stockholm's traditional working-class Södersnack

Stockholm's multiethnolect, the variety of the city's newer working class, sits in juxtaposition to *Södersnack*, the variety of the city's established working class. The relationship between the two varieties is complex. On one hand, multiethnolect has absorbed a large number of slang terms from

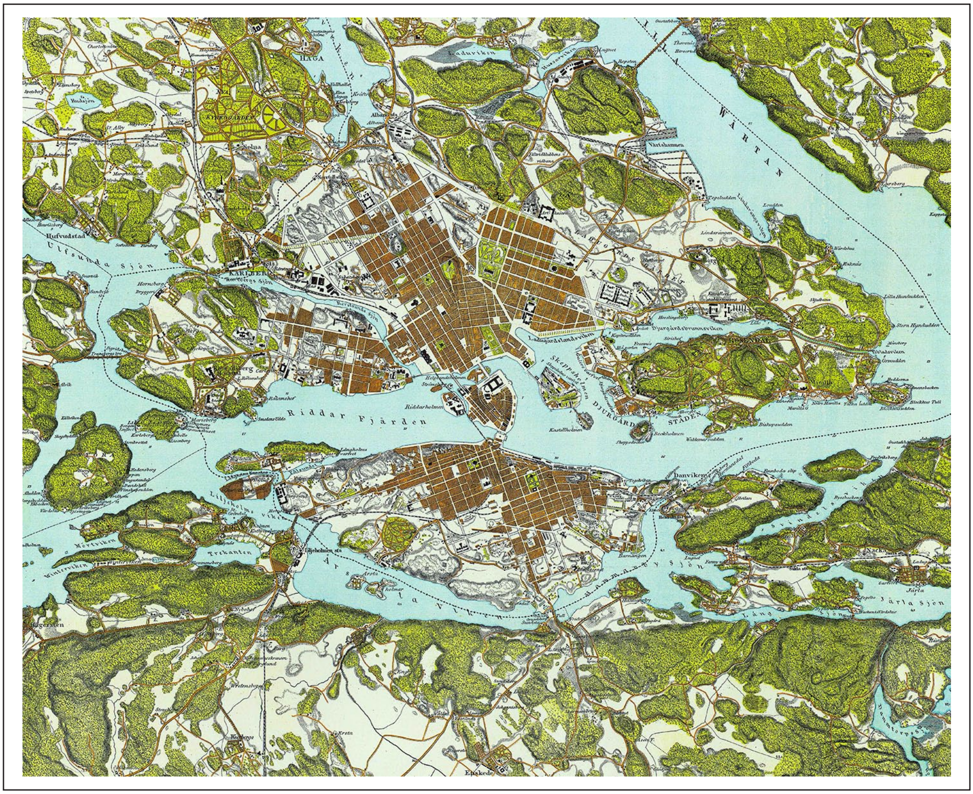


Figure 1. Stockholm in 1841 (Topografiska Corpsen, 1861). Licensed for open use by Stockholm City.

Södersnack (e.g., *beckna* “to sell drugs,” Young, 2018a). On the other hand, its surface phonetics give the impression of a dramatic departure.

Linguistic contact was also rife during Stockholm’s Industrial Revolution, and the massive relocation of rural migrants resulted in intense koinèisation within the speech of their children (Kotsinas, 1988b). Just as is the case for late modernity, the Industrial Revolution was an epoch defined by intense social change and stratification. Aside from the relatively short “Golden Era” of Swedish social democracy (1930s–1980s, Therborn, 1998), Stockholm has long been a hierarchical and stratified place. When industrialization began to partition the citizenry according to their relationship to production, Stockholm’s social classes began settling on separate islands within the city’s dense archipelago. Figure 1 contains a map of the city in 1841. The central island is the original medieval city Gamla Stan, and Södermalm to the south is where the new industrial working class was confined. The growing bourgeoisie spread to Norrmalm and Östermalm towards the North. Since then, the city’s population has grown to fill the full map, but the social classes today continue to be geographically separated.[3] These symbolic and physical enclosures contributed to the emergence and maintenance of Södersnack then (Kotsinas, 1988b; Thesleff, 1912) and of Swedish multiethnolect today.

Stockholm’s traditional industrial working class has undergone a displacement that resembles the Cockney march into Essex (Watt et al., 2014). Starting with the Million Homes Program from 1965 to 1974 (Hall & Vidén, 2005) and accelerated by gentrification on Södermalm, this community has

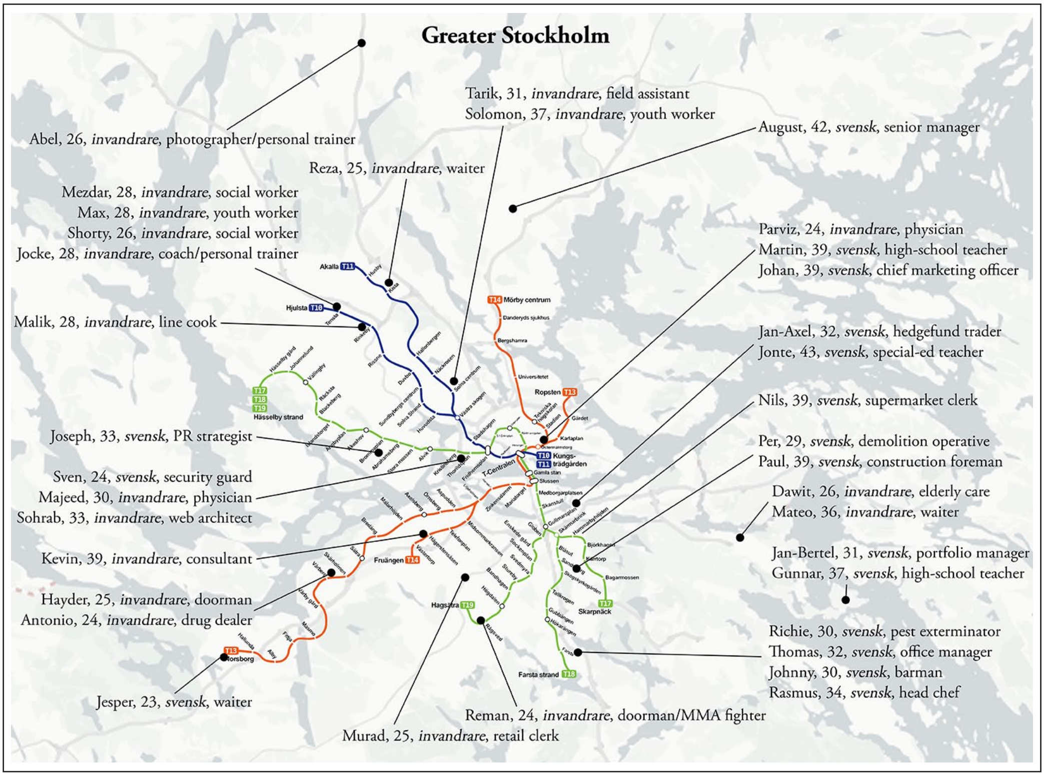

Greater Stockholm
Abel, 26, *invandrare*, photographer/personal trainer
Tarik, 31, *invandrare*, field assistant
Solomon, 37, *invandrare*, youth worker
August, 42, *svensk*, senior manager
Reza, 25, *invandrare*, waiter
Mezdar, 28, *invandrare*, social worker
Max, 28, *invandrare*, youth worker
Shorty, 26, *invandrare*, social worker
Jocke, 28, *invandrare*, coach/personal trainer
Parviz, 24, *invandrare*, physician
Martin, 39, *svensk*, high-school teacher
Johan, 39, *svensk*, chief marketing officer
Malik, 28, *invandrare*, line cook
Jan-Axel, 32, *svensk*, hedgefund trader
Jonte, 43, *svensk*, special-ed teacher
Nils, 39, *svensk*, supermarket clerk
Joseph, 33, *svensk*, PR strategist
Per, 29, *svensk*, demolition operative
Paul, 39, *svensk*, construction foreman
Sven, 24, *svensk*, security guard
Majeed, 30, *invandrare*, physician
Sohrab, 33, *invandrare*, web architect
Dawit, 26, *invandrare*, elderly care
Mateo, 36, *invandrare*, waiter
Kevin, 39, *invandrare*, consultant
Jan-Bertel, 31, *svensk*, portfolio manager
Gunnar, 37, *svensk*, high-school teacher
Hayder, 25, *invandrare*, doorman
Antonio, 24, *invandrare*, drug dealer
Richie, 30, *svensk*, pest exterminator
Thomas, 32, *svensk*, office manager
Johnny, 30, *svensk*, barman
Rasmus, 34, *svensk*, head chef
Jesper, 23, *svensk*, waiter
Reman, 24, *invandrare*, doorman/MMA fighter
Murad, 25, *invandrare*, retail clerk
Akalla T11
Hjulsta T10
Mörby centrum T14
Hässelby strand T17 T18 T19
Ropsten T13
Kungsträdgården T10 T11
T-Centralen
Fruängen T14
Hagsätra T19
Skarpnäck T17
Farsta strand T18
Norsborg T13

Figure 2. Participants matched to their respective neighborhood. Labels provide pseudonym, age, *svensk/invandrare* identification, and occupation. (Map licensed under the Creative Commons and GNU Free Documentation License, Holmér, 2014).

migrated out to the city's periphery. According to popular knowledge and according to fan-tracking for Södermalm's football team *Hammarby* (Young, 2019, p. 29), their migration has appeared to target the suburbs to the south and southeast (see map in Figure 2).

Very little research—sociological or sociolinguistic—has been conducted on this population. Prosody in Södersnack has never been studied, but Öqvist (2010) has proposed that it may have "steeper intonational contours" and a "nasal quality" (p. 256). Three of its long vowels are also reported to be more diphthongal than those in standard Stockholmian (Kotsinas, 1994; Öqvist, 2010), which could translate into more rhythmic contrast than in multiethnolect and the received standard. The vowel in NÄT is merged with LETA (/ɛː / sounds like [i͡eː]) such that both become [i͡eː]. This is known as the *Stockholms e*. The vowel in LAT becomes like LÅS (/ɑː/ sounds like [oᵊː]), and in turn, the vowel in LÅS is shifted upward to become like SOT (/oː/ sounds like [uː]), resulting in a full merge between LÅS and SOT for the most vernacular of speakers. These are catalogued in Table 1, which is an adaptation of Bergman (1946) and Ståhle's (1975) observations reported in Öqvist (2010, p. 254). Kotsinas (1994) is to date the only researcher that has conducted any modern investigation of these vowels. In her study of Swedish youth in the late 1980s, she found that the *Stockholms-e* NÄT vowel was still a thriving variant. Working-class boys produced it 43.7% of the time and working-class girls 17.2%, compared with 4% for upper-class boys and 1% for upper-class girls (p. 331). The backed LAT variant, however, barely showed itself in the data. Boys of both

Table 1. Södersnack vowel trends (adapted from Öqvist 2010, p. 254).

Phonemic description	Phonetic description	Lexical set
/ɛː/ merges with /eː/	[ɛː] merges with [ɪ͡eː]	NÄT merges with LETA
/ɑː/ shifts to /oː/	[ɑː] shifts to [oᵊː]	LAT shifts to LÅS
/oː/ merges with /uː/	[oᵊː] merges with [uː]	LÅS merges with SOT

higher and lower social status produced it only 7% of the time, and girls did not produce any variants (p. 333). She did not investigate LÅS.

2.3 Stockholm's multiethnolect: Europe's "first"

Stockholm's racialized working class has been popularly ascribed a speech variety known as *Rinkeby Swedish* (Kotsinas, 1988a), Europe's earliest-known and Scandinavia's first multiethnolect (Quist, 2012, p. 2). This term *multiethnolect* has been advocated by Clyne (2000) and Quist (2008) to refer to the new urban repertoires spoken by "second generation immigrants" in countries that have experienced significant postwar non-Western migration. Research on these varieties has typically investigated teenage speakers and their construction of distinct and oftentimes oppositional identities (Cheshire, 2013; Cornips & de Rooij, 2013; Doran, 2001; Keim & Knöbl, 2007; Lehtonen, 2011; Pharao et al., 2014; Quist, 2008; Rampton, 2006; Wiese, 2012).

Adult speakers are less commonly discussed in the literature, which is a serious gap. Prior to the present study, I reviewed media discourses on Swedish multiethnolect and found that they extend back as far as 1979 in reference to "semilingual" teens in Stockholm's suburbs (Sveriges Television, 1979). The date of this first report implies that the variety may have been circulating for at least 40 years, so one would expect substantial linguistic focusing in Stockholm by now. Elsewhere in Europe, recent work has started to address the question of adult speakers and to investigate whether these styles now have become full-fledged sociolects (Keim, 2007; Rampton, 2011; Sharma & Rampton, 2015; Sharma & Sankaran, 2011; Young, 2018b). Research on adults was one motivation for Rampton (2011) to disfavor terms like "youth language" while advocating for the designation *contemporary urban vernaculars*. He claims that across Europe, these later-stage linguistic developments share the following three properties—three properties, I would add, that are also very much applicable to Stockholm: (a) they emerged in urban neighborhoods shaped by immigration and class stratification; (b) they are "connected-but-distinct" from migrant languages, the traditional working-class variety, and the standard variety; (c) they are widely known and represented in media and popular culture (p. 291).

The present study seeks to add to this growing genre of literature by investigating the speech of adult male speakers in Stockholm, Sweden. The study focuses on Rampton's (2011) criterion 2 by investigating whether the speech rhythm of Stockholm's multiethnolect indeed is "connected-but-distinct" from both the traditional working-class variety and the standard variety. Additionally, as it pertains to the exceptional longevity of Stockholm's multiethnolect, also discussed above, a second linguistic implication may be at play—that a number of features will have already reached *marker* or even *stereotype* status (Labov, 1972, 2001). This study will investigate whether staccato rhythm might be one such feature.

2.4 Multiethnolects and staccato rhythm

Rhythm, the variable of interest to this paper, is a popular topic when the surface phonetics of multiethnolects are discussed. But only a handful of phonetic studies actually circulate on Swedish multiethnolect. Many more studies have *remarked* that the variety sounds "jerky" or "staccato," even though most of these works are not phonetic (Bijvoet & Fraurud, 2008, p. 22; Bijvoet & Fraurud, 2011, p. 16; Bijvoet & Fraurud, 2016, p. 22; Fraurud, 2003, p. 88; Kotsinas, 1988a, p. 268; Kotsinas, 1990, p. 257; Milani & Jonsson, 2012, p. 54; Nordenstam & Wallin, 2002, p. 257).

Kotsinas offered early impressionistic descriptions of the phonology of Rinkeby Swedish and often referred to its prosody as "jerky" (stötigt) and speculated that it might be due to a reduction in the difference between long and short syllables (Kotsinas, 1988a, p. 268; Kotsinas, 1990, p. 257). This, however, was never tested. Bodén (2007) compared vocalic duration between the received Malmö standard and the local multiethnolect (*Rosengård Swedish*) in a small random sample of her corpus of secondary-school pupils. No significant difference was found in the sample. That is to say, phonologically-long vowels were not shorter and phonologically-short vowels were not longer than those within the standard Malmö Swedish samples (p. 29). She hypothesized that the absence of elisions and reductions might instead have been responsible for the staccato effect that she was perceiving (p. 33).

In the pilot that prefaced this study, I set out to investigate which phonetic features correlated with listener perceptions of "rough" and "non-Swedish" among native-speakers of Stockholm's vernacular (Young, 2018b). Eight short speech samples, all with grammatical and lexical variation removed, were assessed by a panel of 27 native Stockholmers. I then tested which phonetic variants correlated the strongest with their assessments. Low et al.'s (2000) nPVIV of duration (reviewed in Section 4.1) correlated most strongly: the low-nPVIV stimuli garnered assessments of "rough" and "non-Swedish," and the high-nPVIV stimuli garnered assessments of "refined" and "Swedish." The stimuli, however, were not matched guises, so it was not possible to credibly tie these specific evaluations to rhythm, in specific.

Staccato rhythm has also been observed and investigated in other European multiethnolects. Hansen and Pharao (2010) concluded that the staccato impression of Copenhagen's multiethnolect was due to transformations of phonologically long and short vowels. Twelve speakers of multiethnolect and 12 speakers of standard Copenhagen Danish participated in a map task that targeted 33 test words. The authors found that speakers of multiethnolect generally had "equal duration of long and short vowels before syllables containing a full vowel" (p. 93) and that this was generally "due to shortening of long vowels rather than lengthening of short vowels" (p. 91).

Torgersen and Szakay (2012) compared nPVIV of duration (reviewed in Section 4.1) in London's multiethnolect (Multicultural London English, MLE) with that of the white working-class varieties in Havering. They found intervocalic durational contrast to be significantly lower for the former than for the latter. Like Hansen and Pharao (2010), they proposed that the difference might be phonotactic and due to, for example, specific segmental transformations like the monophthongization of the FACE and GOAT vowels (Torgersen & Szakay, 2012, p. 838).

Fagyal (2010), motivated by earlier observations of staccato rhythm in Parisian Verlan (Cerquiglini, 2001; Duez & Casanova, 1997), investigated the speech rhythm of young working-class boys in a Parisian suburb. She divided up her participants by whether they had Algerian heritage or French heritage, yet found no rhythmic significant differences for the three rhythmic measurements taken: (a) percentage of total vocalic duration over total segmental duration, (b) standard deviation of vowel durations, and (c) standard deviation of consonantal durations. One explanation could have been the focus on speaker heritage and L2 effects at the expense of sociolinguistic factors such as self-identification or the possibility that staccato was a shared youth-vernacular feature that spanned all heritage groups (something that I attempt to address in the current study). Another explanation could be that she relied on global rather than local metrics, a topic covered in Section 4.1.

Turning to the segmental phonology of Stockholm's multiethnolect, no conclusive studies have been conducted to date. A number of observations do circulate, however, and I myself have embarked on some earlier work that I will briefly review here. Just as Torgersen and Szakay (2012) proposed for MLE, I have suspected that the long vowels in Stockholm's multiethnolect are more monophthongal. In preliminary work (not peer-reviewed), I have found this to be the case for the vowels in NÄT and LETA, but not for any other vowels (Young, 2019, p. 209). In the same preliminary work, I have also found that unstressed long vowels are qualitatively further from schwa for speakers of multiethnolect than for other speaker groups (Young, 2019, p. 212). As it pertains to quantity, I have found that the difference between long and short vowels in multiethnolect is smaller than in other varieties (Young, 2019, p. 213). However, speakers of multiethnolect also seem to engage in extensive phrase-final lengthening (Young, 2019, pp. 197–198), which would work against low rhythmic alternation. I have also observed that Stockholm's multiethnolect seems to either elide coda rhotics or produce them as flaps whereas the traditional working-class variety usually produces approximants (Young, 2018b, pp. 50–51). Since approximants are both durationally longer and more sonorous than zero realizations and flaps, this could affect intersyllabic contrast—albeit in a way that is unclear. In Section 7.5, I return to these features in order to discuss the present study's findings on rhythm in the context of underlying segments.

2.5 Staccato rhythm and social salience

Fraurud (2003) proposed that, alongside lexicon, prosody appeared to be the most *defining* feature of Stockholm's multiethnolect (p. 87). Bijvoet and Fraurud (2008, 2011) found that laypeople were making similar metapragmatic evaluations. Two listener groups assessed a series of speech stimuli, one of which contained a sample of multiethnic youth language that prompted starkly split assessments. One listener group consisted of monolingual ethnic-Swedish university students; the other listener group consisted of multiethnic secondary-school students from a working-class high school. Whereas most listeners in the first group designated the speech sample as "Rinkeby Swedish" and even pointed to its "staccato-like rhythm," the second group overwhelmingly saw the speech sample as unmarked "good" Stockholm Swedish. The authors note that,

> This speech sample was recorded in a relatively formal situation (a presentation in front of the class). It contains neither slang words nor grammatical deviations and the pronunciation cannot, according to a panel of linguists from Stockholm University, be traced to any particular first language. It is only on the prosodic level that this speech sample diverges from the dominating regional norm—with its light touch of the "staccato intonation" often mentioned in descriptions of multiethnic youth language. (Bijvoet & Fraurud, 2011, p. 16)

Implied here is that "staccato" is salient for outsiders but still not too salient for insiders, which in turn could mean it is in the intermediate stage of evolution, somewhere between what Labov (2001) refers to as *indicator* and *marker* (p. 196).

In a later examination of their data, Bijvoet and Fraurud (2016) found that the impression of "staccato" rhythm for speaker Eleni was sufficient to make listeners think they had heard grammatical errors that were not actually there. They also note that even though Eleni sheds a number of "suburban"[4] features when she style-shifts, she does not quite have access to her prosody.

> Evidently, a single prosodic feature associated with a low-status variety was enough to make listeners also hear what was in reality not there. (Bijvoet & Fraurud, 2016, p. 22)

Prosody and style-shifting emerge as a more explicit topic in Milani and Jonsson's (2012) ethnographic study of youth in a multiethnic suburb of Stockholm. Participant Emre relays an encounter with a police officer whereby he shifts out of his vernacular register into the normative Stockholmian style, shedding his so-called staccato rhythm.

> On the one hand, the performance of "the policeman" is rendered with a lower pitch and an easily recognizable southern inner city Stockholm accent. On the other hand, his own answers are recounted with the "staccato-like" rhythm associated with Rinkeby Swedish. (Milani & Jonsson, 2012, p. 54)

The above accounts imply that (a) some sort of variable is operating in Stockholm's multiethnolect to give the impression of "jerkiness" or "staccato" and (b) that there is some degree of social salience. Testing these two implications constitutes the main purpose of this study. A third aim of this study is to examine the age distribution alongside the style-shifting data to make an apparent-time assessment about the trajectory of speech rhythm in Stockholm.

3 Data collection

3.1 Participants

Thirty-six men, ages 24–43 in 2017, were interviewed for this study between 2015 and 2018. They hail from a stratified sample of social classes and ethnicities that itself constitutes a subsample of a larger ongoing project on the speech of Stockholmers. In order to provide a tangible portrait for the speaker population, they are listed in Figure 2 by pseudonym, age, racialization, and occupation. They are superimposed over a map of Stockholm and visually linked to their home neighborhood. Stockholm's metro line is also included in the map because this is the spatial framework to which the city's residents often associate its social dialects (Bijvoet & Fraurud, 2012). The majority of the working and lower middle-class *invandrare* hail from the northwestern and southwestern suburbs, the majority of the working and lower middle-class *svensk* hail from the southern and southeastern suburbs, and most of the upper middle-class speakers hail from the central four boroughs and the eastern and northern suburbs.

I recruited participants through my own personal and professional networks, and I used the snowball method to seek out new participants through an existing participant's network. Participants received 100 Swedish kronor ($14) per interview; funding was provided by the British Economic and Social Research Council (ESRC).

All participants either have Swedish as a first language or began acquiring it well within the early stages of Lenneberg's (1967) Critical Period. All were born in Sweden except for Antonio (2 yrs), Kevin (4 yrs), and Sohrab (6 yrs), whose age of arrivals are in parentheses; they entered Swedish preschool at age 6. The remaining 33 speakers have a Swedish age of onset, by means of preschool entry or home language, of 0 ($n = 19$), 1 ($n = 8$), 2 ($n = 3$), 3 ($n = 1$), and 4 ($n = 2$). No speaker has a perceivable foreign accent, and all speakers self-reported Swedish as the language they used the most and as either their strongest language or one of their strongest languages. No speaker grew up in monoethnic immigrant enclaves (Stockholm is famously absent of such enclaves; see Aktürk-Drake, 2018), and they all hail from neighborhoods where Swedish is the lingua franca. All are proficient in English[5] and reported English to be a language they used daily—usually via entertainment or social media. Other first languages reported were Egyptian Arabic ($n = 1$), Feyli Kurdish ($n = 1$), Kibembe ($n = 1$), Kurmanji Kurdish ($n = 2$), Luo ($n = 1$), Persian ($n = 4$), Somali ($n = 1$), Swahili ($n = 1$), Spanish ($n = 2$), Tigrinya ($n = 4$), and Turkish ($n = 1$).

Figure 3 contains a population pyramid for Stockholm in 2017 broken down by age and sex. It shows that the age range that I sampled for this study, 24–43 in 2017, is part of the largest two cohorts within the city's population. It also shows that the focus on male speech excludes women,

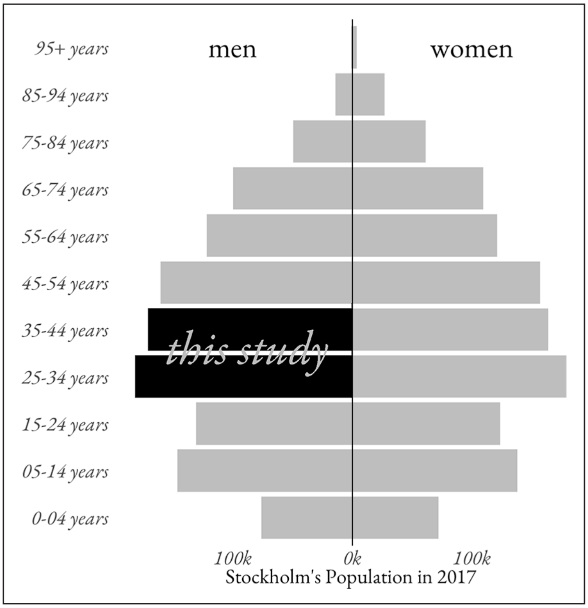
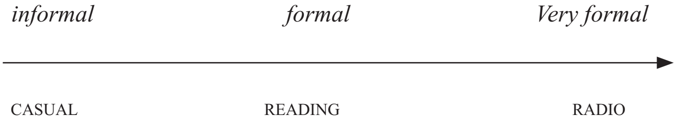


Figure 3. Population pyramid of Stockholm according to age and sex for year 2017 (Statistics Sweden).

who constitute a full *half* of that cohort. Although data collection for women is ongoing and rendering substantial speech data, it remains incomplete. Therefore, this paper will only examine the male speech data collected thus far. As a result of this, it can and will only make claims regarding the portion of Stockholm's population shaded in black in Figure 3.

3.2 Elicitation and interviews

Speakers participated in an adapted *sociolinguistic interview* (Labov, 1972) that consisted of elicited narratives followed by a formal questionnaire followed by two reading tasks. Some speakers also permitted me to record them with their peers; in those cases, I used this data for CASUAL speech data. For speakers who did not participate in peer-group recordings, I used high-involvement narratives from their adapted sociolinguistic interviews for CASUAL speech data. The two reading tasks provided me with so-called READING and RADIO speech data. For READING, I asked the speakers to read aloud a text by the name of *The Circus* (Morris & Zetterman, 2011) with no further instructions. For RADIO, I asked the speakers to read *The Circus* again, albeit while "sounding like an announcer on Radio Sweden."



Morris and Zetterman's (2011) *The Circus* was selected over Engstrand's (1990) *The North Wind and Sun* because its style is less archaic, and I did not want the passage to "key" the voicing of an older speaker. *The Circus* also contains more syllables than *The North Wind and Sun*—367 contra 155—and has multiple occurrences of every Swedish phoneme, whereas *The North Wind* *and Sun* only has single occurrences. The aim of this elicitation approach was to harvest speech data along a formality cline:

The original sample contained 38 speakers, but two were excluded from the analysis due to poor literacy. The remaining 36 speakers had between average and above-average literacy. Extensive social data was also collected in the interviews, which allowed me to place the participants along a numerical social-class scale that is detailed in Section 4.2 and within one of the two racial categories discussed in Section 2.1.

3.3 Recording, transcription, and segmentation

Recordings were made on individual Zoom H1 recorders with self-powered Audio-Technica lavalier microphones. They are in *wav* format, mono, with a sample rate of 16,000 Hertz. The speech data was transcribed by native-language transcribers, which was financed by a grant from the Sven och Dagmar Salén Foundation. The transcriptions were then checked by me and subsequently phonetically time-aligned using SweFA (Young & McGarrah, 2017). I then manually corrected the segmentations in accordance with standard segmentation protocol and the guidelines provided in Engstrand et al. (2000). Segmental metrics were extracted using a customized adaptation of Brato's (2015) script for Praat (Boersma & Weenink, 2017).

The final foot before a pause was included in the calculation (Torgersen & Szakay, 2012; Fuchs, 2016; White & Mattys, 2007; Low et al., 2000; though cf. Thomas & Carter, 2006; Sarmah et al., 2011). I delineated breath groups by pauses that exceeded 150 milliseconds. This is in line with Fuchs (2016, p. 107) but contradicts Thomas and Carter's (2006) recommendation of 70 milliseconds.

In Stockholm Swedish, coda /r/ typically coalesces by means of a sandhi process by which the subsequent /d, n, s, t/ become [ɖ, ɳ, ʂ, ʈ], respectively (Riad, 2014). In front of other consonants, it usually occurs as an approximant[6] or is partially or fully elided. Syllable-final /r/ was included as part of the vowel because the boundaries V+[ɻ] are highly subjective (Thomas & Carter, 2006, p. 341). Syllable-final /ʝ/ was included as part of the preceding vowel for the same reason. On the other hand, intersyllabic /r/ and /ʝ/ (V+/r ʝ/+V), including the external-sandhi effect of coda /ʝ, r/ + onset V (e.g., *han är ung* → *han ä r̲ung*), were kept separate from their surrounding vowels.

Hesitation markers and hesitation lengthening were manually removed as I encountered them. Disfluencies were removed as well. The reading passage contained two words that not all readers knew—"manegen" (circus ring) and "trippelsaltomortal" (triple summersault). These two words were removed from the analysis for all speakers.

The final dataset rendered between 295 and 1517 vocalic intervals per speaker per contextual style, totaling 40,277. The casual dataset contains 21,000 vocalic intervals; reading contains 9704; radio contains 9573.

4 Method

The current study operationalizes rhythm by calculating Low et al.'s (2000) nPVIV with Mishra et al.'s (2012) *energy-F0 integral* (EFI) instead of just duration. In the immediate subsections I offer a review of rhythmic analysis and a basis for why this multidimensional approach was taken.

Following the below review, the analysis makes use of a novel approach that first statistically models how internal factors affect individual rhythmic pairs—without taking into account any social factors (similar to Clopper & Smiljanic, 2015). Additional space is dedicated to detailing how the predictors are coded because the approach is somewhat new, and one goal of the article is to provide a template that is easy to duplicate. Subsequent to modeling the internal predictors, the analysis adds in the social factors to assess their influence on the response variable.

4.1 Contemporary approaches to rhythm

4.1.1 Defining rhythm. Surprisingly few linguistic studies on rhythm are explicit in how they define rhythm. White and Mattys (2007) argue that rhythm "derives from the repetition of elements perceived as similar" (p. 501). Arvaniti (2009) reminds us that White and Mattys' (2007) definition is derived from psychological research (Fraisse, 1963, 1982; Woodrow, 1951) and notes that

> the grouping of stimuli relies not just on duration but on a host of factors, including relative intensity, relative and absolute duration and the temporal spacing of elements (Fraisse, 1963, 1982; Woodrow, 1951). This definition of rhythm implies the presence of meter, which is distinguished from grouping itself: while grouping deals with phenomena that extend over time, meter is an abstract representation that relies on the alternation of strong and weak elements, not on absolute or relative durations (Lerdahl & Jackendoff, 1983). (Arvaniti, 2009, p. 57)

Lerdahl and Jackendoff's (1983) definition of meter, cited above, is especially appealing because it pivots on the notion of local salience. Rather than specifying an internal property, such as *long* versus *short* or *loud* versus *soft*, it focuses on the relativity component: *strong* and *weak*. White et al. (2012) adopt a similar definition for what they call *contrastive rhythm*, a feature that is "evident in any string of sounds in which there is an alternation of strong and weak elements" (p. 665).

In this study, I treat *meter* and *contrastive rhythm* as the same constructions and refer to them as simply *rhythm*, defined as the alternation of strong and weak elements. Although these elements have traditionally been conceptualized as purely durational in the literature, a growing and credible body of work advocates for a multidimensional approach.

4.1.2 Global and local metrics. A common methodological approach that this study avoids is the deployment of *multiple* rhythmic metrics. As I outline below, the rhythm literature has only deployed two algorithms that directly measure the alternation of strong and weak elements. The remaining algorithms are proxy measures that *sometimes* capture rhythmic effects, albeit extraneously.

According to Fuchs (2016, pp. 35–69), contemporary approaches to measuring rhythm can be divided into two categories: *global* and *local*. The most common global metrics originate from Ramus et al. (1999) and are the sum of vocalic intervals divided by the total duration of the sentence (%V), the standard deviation of consonantal durations (ΔC), and standard deviation of vocalic durations (ΔV). These metrics are called global because they assign numerical representations to the overall distribution and variation of segmental matter without specifically modeling sequential contrast. Ramus et al.'s (1999) metrics were the beginning of a distinct era in rhythm studies. Many subsequent studies either incorporated these metrics or iterations of these metrics.

The problem with global metrics is that they serve as proxies; the internal properties of the algorithm were not forged to mathematically simulate rhythmic contrast. So, as proxies, they have a number of inherent weaknesses. Standard deviations of rapid speech are smaller than standard deviations of slow speech, regardless of actual contrastive variation. Consider the standard deviation of sequence [100, 50, 100, 50, 100, 50] versus sequence [50, 25, 50, 25, 50, 25]; both have arguably very similar contrasts with very different standard deviations (27.4 vs. 13.7, respectively). To normalize for rate, *variation coefficients* (the standard deviation divided by the mean) were introduced for consonants by Dellwo (2006) (VarcoC) and for vowels (VarcoV) by White and

Mattys (2007). But variation coefficients (and standard deviations) still do not capture the pivotal *local* component that governs sequential alternation as a construction. Specifically, the sequence [100, 50, 100, 50, 100, 50] renders the same standard deviation and Varco as [100, 100, 100, 50, 50, 50], even though the former sequence contains more sequential contrast than the latter.

Local metrics provide a resolution to this problem because their internal properties actually mathematically model sequential contrast. Two are currently in circulation: (a) the *rhythmic irregularity measure* (RIM, Scott et al., 1986) and (b) the *normalized pairwise variability index* (nPVI, Low et al., 2000[7]) and its alternate variant the *rhythm ratio* (RR, Gibbon & Gut, 2001). For most vowel sequences, the RIM and nPVI/RR produce very similar measurements.[8] Therefore, this study will only use nPVI, and it will take its measurements from vowels. Therefore, I henceforth refer to it as the nPVIV (formula provided in Section 4.2).

The nPVIV algorithm (shown in Section 4.2) is simply a percentage-change calculation with an agnostic denominator. Whereas the percentage change of, say, a temperature rise from 30 degrees to 60 degrees would be $\frac{60-30}{30}$, an agnostic departure point for the change mandates the removal of 30 from the denominator and replacing it with the mean of 30 and 60: $\frac{60-30}{\frac{60+30}{2}}$. This is exactly what the nPVIV does for vowel measurements (see Fuchs, 2014c, for an excellent review). The pairwise index is typically calculated for every vowel pair for a speaker, and then the mean or median is taken to render the nPVIV for that passage. The algorithm is simple and powerful, which has made it popular for all sorts of linguistic investigations of rhythm—typological (Clopper & Smiljanic, 2015; Fuchs, 2016; Gabriel & Kireva, 2014; Grabe & Low, 2002; Prieto et al., 2012; Sarmah et al., 2011), psycholinguistic (Payne et al., 2012; White et al., 2012), and sociolinguistic (Coggshall, 2008; Gilbers et al., 2019; Shousterman, 2015; Thomas & Carter, 2006; Torgersen & Szakay, 2012).

As it pertains to rhythm in European multiethnolects, Torgersen and Szakay (2012) investigated contemporary London varieties on a large scale with a particular focus on Multicultural London English (MLE). They found that young multiethnolectal speakers from Hackney had an overall lower nPVIV, followed by older speakers from Hackney. The highest nPVIV values were produced by the remaining white speakers from both Hackney and Havering (p. 829). The latter point becomes important for the present study's findings on *svensk* working-class speakers (discussed in Section 7.1). In an earlier perception study of Stockholm Swedish (Young, 2018b), I ran a series of regression models that tested the correlation between several phonetic variants and listener assessments of neighborhood and affect. Of all variants tested, speech rhythm as measured by nPVIV showed the strongest statistical correlation to both assessments. These findings motivated the current study and its examination of rhythm in the context of speech production.

There have, however, been some savvy critics of the current rhythmic algorithms. Gibbon (2003) finds the mathematical premise of nPVIV to be problematic. The algorithm assumes a strictly binary interpretation of rhythm and does not have the ability to model unary, ternary dactylic, or anapæstic rhythms. This can therefore result in an averaging out of otherwise important differences. Wiget et al. (2010) caution against the over-reliance of any metric because metrics merely provide approximate indications of stress contrast that often are influenced by extraneous factors (p. 1567). Arvaniti (2009) has expressed similar concern and takes particular exception to studies that apply every algorithm to the data in blanket fashion and then focus on results that fit *a priori* assumptions. To illustrate her point, she applied such a blanket technique to English, German, Greek, Italian, Korean, and Spanish spoken by different speakers and in different speech styles. For most of the algorithms tested, interspeaker and intraspeaker variation was as high *within* each language as across. Arvaniti's (2009) work serves as a reminder that algorithms should only be used if their internal mathematical properties closely align with the natural phenomenon one wishes to model. It also reminds us that measures need to be put in place to control for style and speaker effects, something the present study attempts to do.

4.1.3 Moving beyond the durational paradigm. Durational operationalizations of rhythm have dominated the literature for some time, but a growing body of literature has offered a more multidimensional perspective. Low (1998) incorporated integrals into her early calculations of nPVIV—that is, combined calculations of more than one vowel property. She calculated a duration-F0 integral (duration • mean F0) and an amplitude integral (duration • mean intensity) in her analysis of Singapore English. Fuchs (2016) also incorporated measurements beyond duration in his calculations of nPVIV, including nPVIV of sonority, voicing, F0, the duration-F0 integral, and the duration-amplitude integral in his analysis of Educated Indian English.

Galves et al. (2002) proposed a novel way of investigating rhythm by calculating mean sonority and pairwise sonority contrasts for 25 millisecond-intervals of speech. The authors constructed a function that assigned values close to 1 for the most sonorous intervals and values close to 0 for the most obstruent (p. 324) and applied the function to read-aloud sentences of English, Polish, Dutch, Catalan, Spanish, Italian, French, and Japanese. Results produced measurements for each of the languages that closely reflected their intuitive placement in the "syllable-timed" and "stress-timed" continuum (p. 326).

In her 2010 dissertation, Cumming (2010) found that dynamic F0 contributes to perceiving non-speech sounds and isolated monosyllables as longer than those without a dynamic F0. However, she also found that listeners were more likely to assess stimuli as rhythmic when they had concordance between F0 excursion and duration (2010, p. 191).

Similarly, Fuchs (2014b) conducted a perception experiment whereby listeners assessed stimuli that consisted of a single syllable followed by a second syllable in which the duration of the vowel was varied in 18 steps from 40 to 300 milliseconds and F0 was varied in both syllables at 85, 115, and 145 Hz. Results showed that "for every 60 Hz in F0 difference, there is a 4% increase in perceived duration in the second syllable/vocalic interval." (p. 1951). Fuchs (2014b) then applied this adjustment to the durational measurements taken from each vowel before calculating the nPVIV of duration in his comparative analysis of British and Indian English (p. 1951). He referred to this adjusted algorithm as "perceived variation" and found it to render a difference between the two varieties that was 6.6% higher than the difference that was measured from just duration.

As mentioned above, Low (1998) examined amplitude in her dissertation that prefaced Low et al.'s (2000) seminal study. She calculated the nPVI for root mean square (RMS) amplitude and found higher RMS amplitude nPVI values for British English than for Singapore English (1998, pp. 52–53). He (2012) similarly proposed using amplitude instead of duration for three established measurements—the standard deviation, the variance coefficient, and the nPVIV. This was motivated by a qualitative observation of intensity graphs for L1 English, L1 Mandarin, and L2 English spoken by an L1-Mandarin speaker. All three measurements aligned with the preconceived notion that English was higher-variation than Mandarin and that L2 English spoken by an L1-Mandarin speaker fell in between. Fuchs (2014a) proposed using both average amplitude and duration together in a single calculation by adding individual nPVIV(duration) values to individual nPVIV(amplitude) values and then taking the average for each speaker.

> Crucially, the difference in simultaneous variability in duration and loudness between IndE and BrE was higher than either the difference in variability in loudness or duration. This result shows that a measure of simultaneous variability in duration and loudness, the nPVI-V(dur+avgLoud) suggested in this paper, captures an important aspect of variability in prominence between successive vocalic intervals. (Fuchs, 2014a, p. 292)

This is to say that the integral approach produced a result that was distinct from the separate results on duration and the separate results on amplitude.

The integrated measurements reviewed above have often offered a picture that is much more in alignment with perceptions than durational measurements. Not only do these studies imply that rhythmic contrast is multifaceted, they also show that individual segmental elements often work in synergy such that the presence of one can make listeners think they have heard the other.

4.1.4 Prominence in Swedish. If we return to Lerdahl and Jackendoff's (1983) notion of "strong" and "weak," it becomes necessary to discuss and define prominence. A separate body of literature circulates on prominence alone, unconcerned about questions on rhythm. There is general agreement in this literature that prominence consists of the following three cues: duration, F0, and intensity (Breen et al., 2010; Fry, 1955, 1958; Kochanski et al., 2005; Lieberman, 1960; Turk & Sawusch, 1996; Wagner & Watson, 2010). Researchers, however, disagree on the extent to which each cue contributes to the construction, and this is complicated by the fact that it may differ by language (Wagner & Watson, 2010, p. 925).

As it pertains to Swedish, Fant and Kruckenberg (1994) found duration to be the most important correlate and note that this is unsurprising since Swedish is a quantity language. However, they note also that F0 is of near-equal importance. Fant and Kruckenberg (1994) also observed that intensity typically correlates highly with F0 except at the very peak of an F0 swing, whereby there often appears to be an inverse relationship between energy and F0 (pp. 137–141). In later work, they expanded their study to perception and tested the correlation between assessments of prominence with duration, F0, energy, and subglottal pressure (Fant et al., 2000). While the authors found that all four parameters correlated with prominence in Swedish, they were unable to rank the contributing weight of each (p. 81).

Strangert and Heldner (1995) conducted a perception study that found that "the greater the F0-rise, the stronger the agreement on focus accent. That is, the size of the focus accent cues the degree of prominence" (p. 59). They note, however, that it only partially explains the variation among listener assessments. Heldner and Strangert (1997) later found that "F0-rise is neither necessary nor sufficient for the perception of focus" (p. 55). In subsequent work, Heldner (2003) found that overall intensity and spectral emphasis were strong correlates to focus accents in production.

The literature offers no clear path for prioritizing or assigning weights to duration, intensity, or F0, but what is clear is that they all likely play a key role in constructing prominence in Swedish. Therefore, all three are incorporated in the operationalization of nPVIV, which I detail in the next section.

4.2 Preparing and coding the data for analysis

This study departs from earlier approaches by treating each nPVIV measurement as its own separate observation. Therefore, the analysis will not rely on the correlation of social predictors to *means* or *medians* of nPVIV. Rather, a model will be built that takes into account internal influences on rhythmic contrast, which will facilitate a more credible investigation into whether and to what extent social factors can commandeer rhythm to do their work.

Rhythm is obviously influenced by multiple phonological features, so no examination of external social predictors can be conducted without first identifying language-internal constraints. According to Tagliamonte (2006), any investigation of sociolinguistic variation must first factor in internal predictors (pp. 104–205). For example, in Sharma and Sankaran's (2011) analysis of (t), they included preceding and following segments and word class as predictors. These were modeled alongside gender, age, formality, and so on (p. 412).

To date, only one study has similarly modeled nPVIV at the observation level (Clopper & Smiljanic, 2015), and no study has attempted to account for any additional internal predictors

Table 2. Excerpt of the dataset that a **traditional** nPVI analysis would use. There would be 108 observations (36 speakers · 3 styles), one internal predictor of mean speech rate, the 14 external (social) predictors, and a response variable of median nPVIV.

	style	speaker	Internal Factors	External Factors			Response
			mean_speech_rate	sei_adult	racialization	age	median_npvi_v
1	reading	Abel	228	51	invandrare	26	48.8
2	reading	August	262	73	svensk	42	54.4
3	reading	Antonio	257	11	invandrare	24	45.1
lines 4 through 105							
106	casual	Shorty	195	53	invandrare	26	45.4
107	casual	Tarik	217	51	invandrare	31	42.6
108	casual	Thomas	202	57	svensk	32	44.9

beyond speech rate. The typical procedure is to model social predictors against the speaker's mean or median nPVIV while including mean speech rate as the sole internal parameter (see, e.g., Torgersen & Szakay, 2012, p. 830). This is problematic since a single speaker's nPVIV can move from, for example, 100 to 70 to 55 to 120 to 40 for just a single sequence of vowels. The pair-by-pair nPVI calculations are highly varied and governed by vigorous phonological factors that restrict to what extent a single linguistic community can purpose rhythm for social practice.

Tables 2 and 3 offer a visualization of how the present study differs in its analytic approach. Table 2 shows how my dataset would look in a traditional analysis modeled on means/medians. Table 3 shows my dataset in the new approach, the predictors of which will be discussed in the following subsections on variable coding.

The response variable. Duration, mean F0, and mean intensity were extracted from every vowel. Where F0 could not be measured, the mean value for the speaker for that style was used. Mishra et al.'s (2012) *energy-F0 integral* (EFI) was then calculated followed by nPVIV for each vowel pair.

$$EFI = d_k \cdot energy_k \cdot F0_k \tag{1}$$

where $d_k =$ duration of segment at k^{th} position
where $energy_k =$ mean intensity of segment at k^{th} position
where $F0_k$ = mean fundamental frequency of segment at k^{th} position

$$nPVIV = \frac{\dfrac{|EFI_{n+1} - EFI_n|}{EFI_{n+1} + EFI_n}}{2} \cdot 100 \tag{2}$$

where $EFI =$ energy-F0 integral of the first vowel in the pair
where $EFI_{n+1} =$ energy-F0 integral of the second vowel in the pair

Internal predictors

Phonological quantity, pitch accent, or the lack of accent. Swedish is a pitch-accent language with two lexical accents: *accent 1* and *accent 2*. It is also a quantity language with two phonological categories of vowels, *long* and *short*, that resemble the English or German lax/tense distinction (see Riad, 2014). These interact to form four possible combinations: long vowels with accent 1 (ACCENT_1_LONG), short vowels with accent 1 (ACCENT_1_SHORT), long vowels with accent 2 (ACCENT_2_LONG),

Table 3. Excerpt of the dataset that **this analysis** uses. There are 40,277 total observations (36 speakers · 3 styles • [295 to 1517 observations per speaker]). Random intercepts are speaker and vowel combination. Internal predictors are accent 1 long, accent 1 short, accent 2 long, accent 2 short, unstressed long, unstressed short, phrase-final syllable, coda /r/, speech rate, and mean frequency of constituent lexemes. External social predictors are racialization, social class, and age. The response variable is observation-level nPVIV.

	STYLE	Random effects		Internal Factors									LEXICAL_ FREQ	External Factors	AGE	SOCIAL_ CLASS	Response
		(1| SPEAKER)	(1| VOWEL)	ACCENT_ 1_LONG	ACCENT_ 1_SHORT	ACCENT_ 2_LONG	ACCENT_ 2_SHORT	UNSTRESSED_ LONG	UNSTRESSED_ SHORT	PHRASE_ FINAL	CODA_R	SPEECH_ RATE		RACIALIZATION			NPVIV
1	READING	ABEL	IH$_{1}$_RR_UH$_{0}$	NO	YES	NO	NO	NO	YES	NO	YES	238	3	INVANDRARE	26	25	90.4
2	READING	ABEL	UH0_EH0	NO	NO	NO	NO	NO	YES	NO	NO	156	3	INVANDRARE	26	25	37.3
3	READING	ABEL	EH0_AA0	NO	NO	NO	NO	YES	YES	NO	NO	163	3375	INVANDRARE	26	25	23.7
4	READING	ABEL	AA0_OAH0	NO	NO	NO	NO	YES	YES	NO	NO	160	3401	INVANDRARE	26	25	59.9
5	READING	ABEL	OAH0_AE1	YES	NO	NO	NO	NO	YES	YES	NO	225	54	INVANDRARE	26	25	105.4
6	READING	ABEL	EE3_AH4	NO	NO	NO	NO	YES	YES	NO	NO	151	244	INVANDRARE	26	25	33.3
7	READING	ABEL	AH4_AH0	NO	NO	NO	NO	NO	YES	NO	NO	236	123	INVANDRARE	26	25	19.8
8	READING	ABEL	AH0_AA1	YES	NO	NO	NO	NO	YES	NO	NO	309	1	INVANDRARE	26	25	123.8
9	READING	ABEL	AA1_AEH1	YES	NO	NO	NO	NO	YES	NO	NO	320	3657	INVANDRARE	26	25	115.7
lines 10 through 40,268.																	
40,269	CASUAL	TARIK	EH$_{1}$_AEH$_{0}$	NO	NO	NO	NO	NO	YES	NO	NO	188	545	INVANDRARE	31	73	39.3
40,270	CASUAL	TARIK	AEH0_IH3	NO	NO	NO	YES	NO	YES	NO	NO	204	9708	INVANDRARE	31	73	64.1
40, 271	CASUAL	TARIK	EH4_AH1	NO	YES	NO	NO	NO	YES	NO	NO	203	19605	INVANDRARE	31	73	59.1
40, 272	CASUAL	TARIK	EH1_AE1	YES	NO	NO	NO	NO	YES	NO	NO	237	1698	INVANDRARE	31	73	43.3
40, 273	CASUAL	TARIK	AE1_EH1	YES	NO	NO	NO	NO	YES	YES	NO	293	4519	INVANDRARE	31	73	46.7
40, 274	CASUAL	TARIK	UH0_EE2	NO	NO	NO	NO	YES	YES	NO	NO	225	1	INVANDRARE	31	73	96.8
40, 275	CASUAL	TARIK	EE2_AH4	NO	NO	NO	NO	YES	YES	NO	NO	167	1	INVANDRARE	31	73	39.6
40, 276	CASUAL	TARIK	AH4_EH0	NO	NO	NO	NO	NO	YES	NO	NO	180	1	INVANDRARE	31	73	46.8
40, 277	CASUAL	TARIK	EH0_YH0	NO	NO	NO	NO	NO	YES	NO	NO	182	41	INVANDRARE	31	73	34.9

and short vowels with accent 2 (accent_2_short). Each vowel in the dataset was coded according to acoustic prominence, the exact procedure of which is detailed in Young (2019, pp. 129–133).

Table 3 offers an example for how the coding is distributed among the four accent/quantity combinations. For example, if either vowel in the nPVIV pair is prominent, phonologically long, and part of an accent-1 word, then the predictor accent_1_long is coded yes as is the case for the pairs that contain AE1 [^{1}ɛː] on lines 5, 40,272, and 40,273 and for the pairs that contain AA1 [^{1}ɑː] on lines 8 and 9.

A lack of prominence strips a word of its lexical pitch accent. For example, on line 6 of Table 3, the 3 in EE3 ([^{2}eː]) is the SweFA code for stress accent 2. The 4 in AH4 ([a$_{2}$]) is the SweFA code for the post-tonic accent following stress accent 2. They are both acoustically non-prominent in this occurrence, so their lexical accents are stripped, and they become [$^{2\to0}$e] and [a$_{2\to0}$], respectively. If one of the two vowels in the nPVIV pair lacks acoustic prominence and is phonologically long, then the predictor unstressed_long is coded yes as is the case for AA0 [ɑː$_{0}$] on lines 3 and 4, EE3 [$^{2\to0}$eː] on line 6, and EE 2 [$^{2\to0}$eː] on lines 40,274 and 40,275. If one of the vowels in the nPVIV pair lacks acoustic prominence and is phonologically short, then the predictor unstressed_short is coded yes. This is the case for all of the vowel pairs in the snapshot of the dataset (lines 1–9 and lines 40,269–40,277).

Phrase-finality. According to Klatt (1976), English vowels lengthen within a pre-pausal foot. This is why Thomas and Carter (2006) and Sarmah et al. (2011) chose to omit phrase-final vowels from their calculation of nPVIV. On the other hand, Low et al. (2000), White and Mattys (2007), Torgersen and Szakay (2012), and Fuchs (2016) chose to include them. Fuchs (2016) examined the effects of phrase finality on nPVIV and found that its inclusion or exclusion contributed to negligible change for British English. However, he found that including phrase-final syllables in his analysis of Indian English resulted in a slight decrease in nPVIV for read-aloud speech and a slight increase in nPVIV for spontaneous speech (p. 103). White and Mattys (2007) make a strong case for inclusion: first, lengthening does not occur before all pauses, and second, if there is lengthening, it may contribute to the overall perception of rhythm (p. 507). Since preliminary work has identified phrase-final lengthening as a potential feature of Stockholm's multiethnolect (Young, 2019, pp. 197–198), it is even more important to code for it (or remove it from the analysis). Therefore, every vowel that falls before a pause is coded yes for the id phrase_final. A pause is defined as any break of 150 milliseconds or more. Lines 5 and 40,273 in Table 3 exemplify two phrase-final vowels.

Syllable-final /r/. As discussed in Section 3.3, I take the approach of Thomas and Carter (2006) and combine syllable-final /r/ with the preceding vowel, treating them as a single segment. Since, however, this renders larger vocalic measurements, which, in turn, renders larger nPVIV measurements, they must be accounted for in the statistical model. These occurrences are coded yes in the model under the heading coda_r. Line 1 in Table 3 contains an example.

Speech rate. Speech rate has been shown to affect rhythm *even* in cases where rhythm is operationalized by means of a *rate-normalizing* algorithm. The literature is consistent in showing that higher rates of speech result in generally less durational contrast (Dellwo, 2010; Deterding, 2001; Torgersen & Szakay, 2012; although cf. Fuchs, 2016, pp. 150–151). In the case of London's multiethnolect, Torgersen and Szakay (2012) have demonstrated a correlation between higher rates of speech and lower nPVIV (pp. 831–832). This is also the case with the present dataset. The relationship is nonlinear and strongest at higher speech rates, so the natural logarithm is used in the statistical model. Rate is measured as the mean duration of the two syllables that contain each vowel pair (Deterding, 2001; Thomas & Carter, 2006; though c.f. Torgersen & Szakay, 2012, p. 828, and Dellwo, 2010, p. 5).[9]

Table 4. Correlation coefficients *R* between the three social predictors.

	race	class	age
race	1.0	–0.4	–0.5
class		1.0	0.4
age			1.0

Lexical frequency. Lexical frequency can affect vowel reduction (Pluymaekers et al., 2005) and fuller realizations of segments (Zhao & Jurafsky, 2009), both of which can affect intervocalic contrast. Naturally, the READING and RADIO styles control for lexical frequency because all participants read the same passage. For CASUAL speech, however, lexical inventory will vary. Frequency data was harvested from the *Swedish Spoken Language Corpus* (SSLC) (Allwood, 1999) and coded for each word. Despite being a modest corpus of 1.2 million words, the SSLC is the largest spoken corpus of Swedish that I am aware of. Frequency was coded for each vowel pair in the column headed LEXICAL_FREQ (Table 3). Words missing from the SSLC were given a frequency of 1, because the speaker must have heard the word at least once in order to produce it. Vowel pairs that straddle words take the mean frequency of those two words. Since lexical frequency is Zipf-distributed, its natural logarithm is used in the model (Pluymaekers et al., 2005; Zhao & Jurafsky, 2009).

External predictors

Contextual style. Each vowel pair is coded under the STYLE predictor as CASUAL, READING, or RADIO. The reference factor in the model is CASUAL since that is the baseline style.

Racialization. As reviewed in Section 2.1, speakers colloquially refer to themselves as *svensk* or *invandrare*. The two emic designations are coded as binary factors for the RACIALIZATION predictor. The reference factor in the model is *svensk* since that is the hegemonic group.

Social class. Inspired by much of the early First Wave variationist work (Fontanella de Weinberg, 1974; Labov, 2001, 2006 [1966]; Trudgill, 1974; Wolfram, 1969), a 100-point index was devised for the CLASS predictor from the first principal component of a principal components analysis (PCA) of six social factors: (a) occupational status according to Ganzeboom and Treiman's (2003, 2018) International Socio-Economic Index of occupational status ISEI, (b) income, (c) formal education, (d) parental occupational status according to ISEI, (e) parental formal education, and (f) taste for 60 consumer items from Experian Ltd and InsightOne Nordic AB's (2013) market segmentation of Stockholm. Details on the index are located in Young (2019, pp. 144–152).

Age. The AGE predictor is coded as the age of each participant in 2017. This is preferable to birth year, which would force a 2000-point scale onto the model.

Orthogonality of social predictors. Participants were sampled during the fieldwork such that the three social predictors maintained a correlation coefficient of 0.5 or less between one another. Table 4 provides a matrix showing the correlation coefficients between the three predictors. The predictors show a middle range of orthogonality, which implies that variance inflation factors should especially be heeded when the predictors are included in the same statistical models.

Random effects

Speaker. For this dataset, speaker is a particularly important random effect because the number of observations per speaker varies. The lowest observation count is $n = 295$ for August, and the highest is $n = 1517$ for Jocke. In a simple linear-regression model, the social factors associated with Jocke would therefore weigh five times more than those associated with August. In a mixed-effects model, random effects will mathematically adjust for this bias.

Vowel. Every observation-level nPVIV consists of a vowel pair. These pairs are not unique for every observation; rather, there are 3334 unique pairs within the 40,277 nPVIV calculations.

5 Analysis

Three multiple regression analyses were run. The first model includes just internal predictors, the second includes all external predictors, and the third includes an optimal internal/external model. Each model output is provided in Table 5. For the internal model, the following call was used in R:

> npviv ~ accent_ 1_ long + accent_ 1_ short + accent_ 2_ long + accent_ 2_ short + unstressed_ long + unstressed_ short + pair_ coda_ r + phrase_ final + *log* (lexical_ freq) + *log* (speech_ rate) + (1|speaker) + (1|vowel).

Since the main research question concerns the speech of young racialized working-class speakers, these three combinations are possible when adding social factors to Model 1:[10]

$$\begin{aligned} \rightarrow age + race : class \\ age : race + class \\ age : class + race \end{aligned}$$

The general rule of thumb in a regression model is ten participants per parameter, and according to Heo and Leon (2010), a fourfold increase is needed in order to run three-way interactions ($n = 80$). Since the sample size of the current dataset is 36, I ran a separate mixed-effects linear regression model for each of the three combinations, and the $age + race : class$ model showed the best fit. The other two models ($age : race + class$ and $age : class + race$) had higher AIC and BIC values and also rendered unacceptably high variance inflation factors.

The $age + race : class$ paradigm was therefore selected for Model 2 in Table 5.[11] Contextual style was then added in interaction with each of the social predictors (style* age + style* racialization* social_ class). The random slope of speaker and style (1+style|speaker) and vowel and style (1+style|vowel) were also included; even if speakers respond as a group to the interview conditions in a consistent manner, a portion of the response will inevitably be subject to individual variation. Lower literacy is an example of one such random effect that only applies to certain experimental conditions (the reading task) but not to others (peer-group recordings). Such conditional effects are captured with a random slope (Barr et al., 2013; Johnson, 2014, p. 18). The following call was used in R for Model 2:

> npviv ~ accent_ 1_ long + accent_ 1_ short + accent_ 2_ long + accent_ 2_ short + unstressed_ long + unstressed_ short + pair_ coda_ r + phrase_ final + *log* (lexical_ freq) + *log* (speech_ rate) + style*age + style*racialization*social_ class + (1+style|speaker) + (1+style|vowel)

Table 5 Model 2 contains the output, but the number of insignificant predictors—as well as the mildly higher VIF for style (5.55) and age:style (5.46)—suggests that it might not be the best fit for the data. Therefore, I removed the style interaction in a stepwise fashion from every other external predictor and qualitatively assessed the model response. Model 3 in Table 5 was the optimal

Table 5. Mixed-effects linear regression models with nPVIV as the response variable. Model 1 includes internal predictors only. Model 2 includes all internal and external predictors and all possible interactions. Model 3 is the optimal model for internal and external predictors and interactions. For categorical predictors, the reference category is in italics (e.g., *yes*). Coefficients are indicated in the center column, standard errors in the parentheses, and variance inflation factors (VIF) to the right.

	Model 1		**Model 2**		**Model 3**	
	Internal only		*Internal & all external*		*Optimal model internal & external*	
RESPONSE VARIABLE	**nPVIV$_{EFI}$**		**nPVIV$_{EFI}$**		**nPVIV$_{EFI}$**	
(Intercept)	−3.45 (3.00)		−4.57 (5.65)		−4.72 (5.22)	
INTERNAL PREDICTORS		VIF		VIF		VIF
ACCENT_1_LONG·*yes*	27.01 (0.71)***	1.64	26.21 (0.70)***	1.25	26.21 (0.70)***	1.25
ACCENT_1_SHORT·*yes*	18.68 (0.70)***	1.19	18.24 (0.70)***	1.09	18.24 (0.70)***	1.09
ACCENT_2_LONG·*yes*	23.83 (0.92)***	1.36	23.55 (0.91)***	1.15	23.54 (0.91)***	1.15
ACCENT_2_SHORT·*yes*	14.88 (0.68)***	1.18	14.46 (0.68)***	1.09	14.46 (0.68)***	1.09
UNSTRESSED_LONG·*yes*	8.05 (0.70)***	2.11	8.20 (0.67)***	1.41	8.20 (0.67)***	1.41
UNSTRESSED_SHORT·*yes*	16.62 (0.68)***	1.53	16.09 (0.67)***	1.24	16.09 (0.67)***	1.24
CODA_R·*yes*	7.92 (1.00)***	1.02	7.86 (0.97)***	1.01	7.86 (0.97)***	1.01
PHRASE_FINAL·*yes*	−1.16 (0.64)$^{\circ}$	1.04	−1.29 (0.66)*	1.02	−1.29 (0.66)*	1.02
log (LEXICAL_FREQ)	−0.03 (0.07)	1.08	−0.02 (0.08)	1.04	−0.02 (0.08)	1.04
log (SPEECH_RATE)	6.41 (0.53)***	1.12	5.91 (0.53)***	1.05	5.91 (0.53)***	1.05
STYLE ALONE PREDICTORS				VIF		VIF
STYLE·*reading*			−5.42 (4.69)	5.55	−6.44 (2.57)*	3.02
STYLE·*radio*			−5.89 (4.94)	5.55	−5.70 (2.70)*	3.02
SOCIAL PREDICTORS				VIF		VIF
AGE			0.38 (0.15)**	1.44	0.39 (0.13)**	1.23
RACIALIZATION·*invandrare*			−12.19 (3.07)***	2.59	−12.19 (3.05)***	2.58
CLASS			−0.09 (0.04)*	2.07	−0.09 (0.04)*	2.02
RACIALIZATION·*invandrare*:CLASS			0.15 (0.05)**	2.39	0.15 (0.05)**	2.37
STYLE & SOCIAL INTERACTION PREDICTORS				VIF		VIF
STYLE·*reading*:AGE			−0.04 (0.14)	5.46		
STYLE·*reading*:RACIALIZATION·*invandrare*			5.88 (2.95)*	3.06	5.89 (2.90)*	3.05
STYLE·*reading*:CLASS			0.12 (0.04)**	3.02	0.12 (0.04)**	2.79
STYLE·*reading*:RACIALIZATION·*invandrare*:CLASS			−0.10 (0.05)$^{\circ}$	2.52	−0.10 (0.05)$^{\circ}$	2.47
STYLE·*radio*:AGE			0.01 (0.15)	5.46		
STYLE·*radio*:RACIALIZATION·*invandrare*			7.71 (3.09)*	3.06	7.71 (3.04)*	3.05
STYLE·*radio*:CLASS			0.09 (0.04)*	3.02	0.09 (0.04)*	2.79
STYLE·*radio*:RACIALIZATION·*invandrare*:CLASS			0.09 (0.06)	2.52	0.09 (0.05)	2.47
AIC	396 187.13		395 675.23		395 666.78	
BIC	396 307.57		396 002.16		395 976.51	
Log Likelihood	−198 079.56		−197 799.61		−197 797.39	
Num. obs.	40 277		40 277		40 277	
Num. groups: VOWEL	3334		3334		3334	
Num. groups: SPEAKER	36		36		36	
Var: VOWEL (Intercept)	167.72		102.72		102.73	
Var: SPEAKER (Intercept)	26.11		15.40		15.24	

(Continued)

Table 5. (Continued)

	Model 1 *Internal only*	**Model 2** *Internal & all external*	**Model 3** *Optimal model internal & external*
Var: Residual	1 036.32	1 019.33	1 019.34
Var: VOWEL STYLE·*reading*		207.99	207.80
Var: VOWEL STYLE·*radio*		245.55	245.32
Cov: VOWEL (Intercept) STYLE·*reading*		−21.34	−21.28
Cov: VOWEL (Intercept) STYLE·*radio*		−24.47	−24.35
Cov: VOWEL STYLE·*reading* STYLE·*radio*		224.76	224.56
Var: SPEAKER STYLE·*reading*		10.17	9.69
Var: SPEAKER STYLE·*radio*		11.78	11.21
Cov: SPEAKER (Intercept) STYLE·*reading*		−3.38	−3.20
Cov: SPEAKER (Intercept) STYLE·*radio*		−7.22	−6.93
Cov: SPEAKER STYLE·*reading* STYLE·*radio*		8.96	8.59

***$p < 0.001$, **$p < 0.01$, *$p < 0.05$, °$p < 0.1$.

Table 6. Calculation of a *base constant* for a "typical" vowel pair, as predicted by Model 1 in Table 5. The assumption is that the word frequency is 7000, speech rate is 150 milliseconds, and that phrase-finality and an /r/ in coda position are not present. This *base constant* is added to the *accent-length coefficient* in Table 7 in order to demonstrate the internal effects that accent (or the lack thereof) has on nPVIV in the dataset.

Estimated nPVIV for a typical vowel pair within a typical word at a typical speech rate:
= [*base constant*] + [*accent-length coefficient*]
= *Intercept* + *coefficient*(*lexical_freq*) + *coefficient*(*speech_rate*) + [*accent-length coefficient*]
= [−3.5] + [−0.03 · *ln*(7000)] + [6.4 · *ln*(150)] + [*accent-length coefficient*]
= [−3.5] + [−0.3] + [32.1] + [*accent-length coefficient*]
= 28.3 + [*accent-length coefficient*]

model when AIC, BIC, log likelihood, VIF, and number of significant predictors were assessed in toto. It contains the same call as in Model 2, albeit with the interaction STYLE* AGE removed.

6 Results

6.1 Internal influences on rhythm

Model 1 in Table 5 predicts that all internal factors have a significant effect on nPVIV. Tables 6 and 7 offer a calculation of a case study of how each phonological type affects rhythm. Table 6 contains the calculation for a base constant for the rhythm of a vowel pair that has a word frequency of 7000 (a middle-range figure) and an average speech rate of 150 milliseconds (a middle-range rate). From this base constant, estimated nPVIV values are calculated for each of the main vowel combinations in Table 7.

Table 7 aims to show to what extent rhythm is internally-governed, absent of social effects. In other words, the internally-constrained *envelope of variation* (Labov, 1972) is substantial and lies between approximately 36.4 on the lower end and 71.9 on the upper end. Clear here is that rhythmic

Table 7. Vowel-pair combinations in dataset rank-ordered by effect on nPVIV (highest to lowest), as predicted by Model 1 in Table 5.

Stress-length combinations within pairs	Accent-length coefficient		Estimated nPVIV				*n* in dataset	% *n* in dataset
accent-1 long V + unstressed short V	27.0+16.6=	43.6	+28.3=	71.9	high ↑ alternation ↓ low	less ↑ staccato ↓ more	4,616	11.5
accent-2 long V + unstressed short V	23.8+16.6=	40.4	+28.3=	68.7			2,880	7.2
accent-1 short V + unstressed short V	18.7+16.6=	35.3	+28.3=	63.6			2,550	6.3
accent-1 long V + unstressed long V	27.0+8.1=	35.1	+28.3=	63.4			1,901	4.7
accent-2 long V + unstressed long V	23.8+8.1=	31.9	+28.3=	60.2			520	1.3
accent-2 short V + unstressed short V	14.9+16.6=	31.5	+28.3=	59.8			3,248	8.1
accent-1 short V + unstressed long V	18.7+8.1=	26.8	+28.3=	55.1			1,092	2.7
unstressed short V + unstressed long V	16.6+8.1=	24.7	+28.3=	53.0			7,707	19.1
accent-2 short V + unstressed long V	14.9+8.1=	23.0	+28.3=	51.3			916	2.3
unstressed short V + unstressed short V	16.6=	16.6	+28.3=	44.9			11,537	28.6
unstressed long V + unstressed long V	8.1=	8.1	+28.3=	36.4			2,236	5.6
remaining combinations		*n/a*		*n/a*			1,074	2.6
					TOTAL:		40,277	100

variation is dominated by the distribution of prominence and phonological quantity, two features that are highly language-specific. This explains why the internal coefficients dwarf the external coefficients in size and continue to retain both strength and significance in Models 2 and 3.

6.2 External influences on rhythm

6.2.1 Age. For all speech styles, Model 3 predicts that age will have a main effect such that for every year older a speaker is, his 100-point nPVIV will increase by 0.39. This means that if all other factors remain the same, a 22-year-old speaker will have an nPVIV for an accent-1 long vowel adjacent to an unstressed short vowel that resembles the nPVIV of a 44-year-old speaker producing an accent-1 short vowel adjacent to an unstressed short vowel. Referring back to Table 7, the difference between the former and the latter is 8.5 ($71.9 - 63.6 = 8.3$), which is the approximate effect of age over a 22-year span ($22 \cdot 0.39 = 8.58$).

6.2.2 Racialization and social class. For casual speech, Model 3 predicts that *invandrare* speakers will have a lower nPVIV than *svensk* speakers. But this is complicated by social class. Racialization only has a minor effect on nPVIV when social class is high. When social class is low, it has a large effect. An elite *invandrare* man with a social class index of 100 is predicted to have an nPVIV that is 2.81 higher than his *svensk* counterpart ($100 \cdot 0.15 - 12.19 = 2.81$). On the other hand, a lower-class *invandrare* man with a social class index of 1 is predicted to have an nPVIV that is 12.04 lower than his *svensk* counterpart ($1 \cdot 0.15 - 12.19 = -12.04$). This is strikingly high and implies that among the lower classes in Stockholm, *invandrare* men are producing a contrast effect for an accent-1 long vowel adjacent to an unstressed short vowel that resembles an accent-2 short vowel adjacent to an unstressed short vowel produced by his *svensk* counterpart. Referring back to Table 7, the difference between the former and the latter is 12.1 ($71.9 - 59.8 = 12.1$).

Important also is that elite speakers, regardless of racialization, are predicted to have an nPVIV between the two lower-class groups. A lower class *svensk* speaker is predicted to have an nPVIV that is 8.91 *higher* than a higher-class *svensk* speaker ($(100-1) \cdot -0.09 = -8.91$). A lower-class *invandrare* speaker is predicted to have an nPVIV that is 5.94 *lower* than a higher-class *invandrare* speaker ($(100-1) \cdot (0.15-0.09) = 5.94$).

6.2.3 Contextual style. When comparing the main effects for READING and RADIO between Models 2 and 3, one sees that the significance level is sensitive to the interaction load. In Model 2, the main effects are not significant, but in Model 3 they are. The complicated interaction of RACIALIZATION and CLASS illustrates why the main effect would be so fragile: lower-class speakers are scattered both below and above elite speakers, so any working-class style-shift that targeted elite speech would not be unidirectional. Lower-class *svensk* speakers are predicted to move to a lower nPVIV when shifting to READING and RADIO styles, decreasing their nPVIV by 6.32 and 5.61, respectively ($-6.44 + 1 \cdot 0.12$ and $-5.70 + 1 \cdot 0.09$, respectively). The lowest-class *invandrare* speakers are not predicted to move much at all when shifting to READING style, but are predicted to increase their nPVIV by 2.01 when shifting to RADIO style ($-6.44 + 5.89 + 1 \cdot 0.12 + 1 \cdot -0.10$ and $-5.70 + 7.71 + 1 \cdot 0.09 + 1 \cdot -0.09$, respectively).

Both higher-class *svensk* and *invandrare* speakers, however, are predicted to style-shift in the same direction. Higher-class *svensk* men are predicted to increase their nPVIV by 5.56 and 3.3 when shifting to READING and RADIO styles, respectively ($-6.44 + 100 \cdot 0.12$ and $-5.70 + 100 \cdot 0.09$, respectively). Higher-class *invandrare* men are predicted to increase their nPVIV by 1.45 and 2.01 when shifting to READING and RADIO styles, respectively ($-6.44 + 5.89 + 100 \cdot 0.12 + 100 \cdot -0.10$ and $-5.70 + 7.71 + 100 \cdot 0.09 + 100 \cdot -0.09$, respectively).

6.2.4 A systematic overview of the results. While individual coefficients do offer insight, they are by necessity stripped of context and comparative utility. Figure 4 therefore contains a case study of the implications of Table 5, Model 3. Four speakers from opposing ends of the class and race spectrum are modeled: lower-class and higher-class *svensk* and lower-class and higher-class *invandrare*. A class index of 1 is used for lower class, and a class index of 100 is used for higher class. A fifth speaker, an *invandrare* within the lower-middle range of the class distribution, is also provided (class index = 30). Speakers on the lower end of the spectrum are referred to as working-class (WC); speakers on the upper end as upper middle-class (UMC), and the centrally-located speaker as lower middle-class (LMC). Age in 2017 is modeled at three intervals: 40 (born in 1977), 30 (born in 1987), and 20 (born in 1997).

Overall, nPVIV decreases in apparent time, independent of other social factors. However, the interaction with race and class enriches this picture. Although the entire system appears to be moving towards low alternation, the social stratification between *invandrare* and *svensk* remains stratified relative to the speech community. Among lower-class *invandrare* men, staccato rhythm is a consistent variant that becomes more and more staccato from 1977 to 1997 while maintaining its relative difference from the overall system. For lower-class *svensk* men, the high-alternation variant also appears to becoming lower and lower, albeit in consistent relationship to the overall movement of the system.

The interaction with style adds another dimension that is of particular importance to the question of rhythm and social salience. As the system moves in apparent time toward an overall lower degree of alternation, speakers continue to move toward a shared rhythmic norm in more formal styles. The CASUAL (*informal*) recordings have clear social variation, and the unprompted READING task (*formal*) shows some sort of convergence towards a norm, albeit inconsistently (discussed further in Section 7.3). The RADIO style (*very formal*), however, shows the most consistent convergence toward a shared norm.

7 Discussion

7.1 Race and class stratification

The findings indicate that rhythm is socially stratified such that it splits three ways in the vernacular of men: low "staccato" alternation within the racialized *invandrare* working class, high alternation within the *svensk* working class, and intermediate alternation within the middle classes/elites. The latter finding supports much of the earlier local work that has described the prosody of Stockholm's multiethnolect as "staccato." Beyond local relevance and of interest to sociolinguistic theory is the finding that the main stratification is *horizontal* rather than the top-to-bottom trend seen in typical class-based sociolectal variation:

	invandrare lower class	higher class	*svensk* lower class
mean nPVIV	53.5	← 61.3 →	69.3

This finding is somewhat surprising in light of some proposals that European multiethnolects have close ties to their cities' respective traditional working-class varieties. For example, Rampton (2011) has found that speakers of London's contemporary urban vernacular also use "a significant number of traditional London vernacular features" (p. 288) such as TH-dropping and fronting, H-dropping, alveolar ING, and centralized MOUTH (p. 284). Doran (2001) has described Parisian Verlan as featuring both foreign loans and archaic blue-collar French slang (p. 98). Similarly, in my

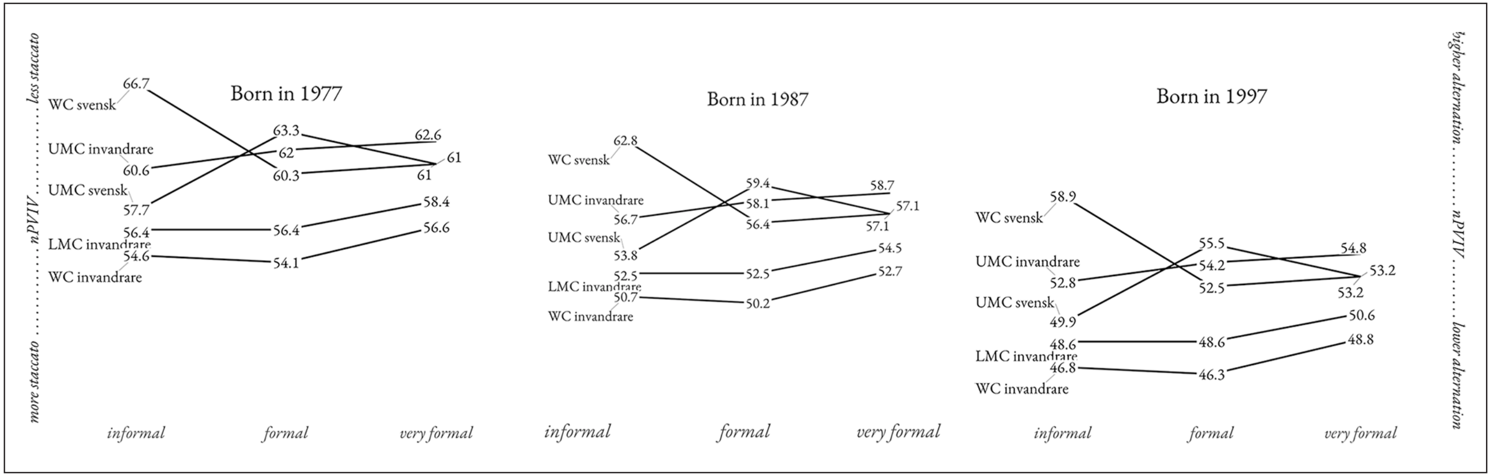


Figure 4. Rhythm, class, and race in apparent time for Stockholm, year 2017: A simulation built from the coefficients in Table 5, Model 3 for five hypothetical speakers in three contextual styles from three cohorts—born in 1977, 1987, 1997. Hypothetical speakers are working-class (WC) *svensk*, working-class (WC) *invandrare*, lower middle-class (LMC) *invandrare*, upper-middle class (UMC) *svensk*, upper-class (UMC) *invandrare*. Working class is modeled with a class index of 1; lower-middle class is modeled with 30; upper-middle class is modeled with 100. Contextual styles are *informal, formal* and *very formal.*

own research on loanwords in Stockholm's multiethnolect (Young, in press, 2021), a large number of older Södersnack words seem to have been re-appropriated into the contemporary vernacular alongside foreign loans.

In light of other work, these results are less surprising. As was discussed in Section 4.1, Torgersen and Szakay (2012) found that while young speakers of MLE had the lowest overall nPVIV, the highest nPVIV values were for the white working-class speakers from Hackney and Havering. Similarly, in an investigation of the FACE and PRICE vowels in a multiethnic school in East London, Gates (2018) found that all students of color were partaking in the multiethnolectal monophthongization to some degree. The glaring exception was White British girls who, despite being a significant minority in their multiethnic school, produced conservative Cockney diphthongs for FACE and PRICE. She proposed that "they linguistically and socially construct Whiteness in a way not often found in multicultural, ethnically-diverse communities" (p. 47).

In Copenhagen, a city similar to Stockholm in terms of language and sociocultural context, one variable also seems to have dual working-class variants that straddle an "intermediate" middle-class feature. Strong t-affrication [ʦ] is a traditional Low Copenhagen variant for /t/ (Quist, 2010, p. 638) and is still a widely-known stereotype of the "Amager" working-class persona. On the other hand, mild affrication [t^{s}] is the standard middle-class (and traditional High Copenhagen) variant. In recent years, however, palatalized [t^{j}]—a variant produced at the other end of the articulatory range—has emerged as the multiethnolectal variant (Pharao & Maegaard, 2017) in some of the very neighborhoods where [ʦ] had dominated earlier.

In Stockholm, as is the case in Copenhagen, London, and Paris, the racialized working class lives in much closer proximity to the white working class than it does to the middle classes. Therefore, results like these implicate two possible catalysts—one being "strong mechanical" and the other being "strong identity" in nature. Within the "strong mechanical" hypothesis, the group second language-acquisition effect is so strong that its output remains impermeable to the locally-available variant. So if a substrate variant like low-alternation prosody dominates the feature pool in a multiethnic housing estate and overwhelms the inputs that subsequent generations in that housing estate are exposed to (by virtue of many heritage languages having such prosody), it will emerge despite the fact that high-alternation prosody is the dominant feature in the greater urban area outside of that housing estate. Within the "strong identity" hypothesis, children in the housing estate perceive from a very early age which social types produce which variants, and a norm is established during adolescence that involves an identity move toward one social type (e.g., working-class immigrant) and away from another (e.g., working-class Swede). So if a substrate variant like low-alternation prosody is competing against other variants within the feature pool, its chance for domination and selection into the new vernacular increases if its phonetic "opposite" is already in use by an opposing group.

The present study does not have the data to prefer one hypothesis over the other, but I believe that discussing these results, and similar results elsewhere, within this schematic can be beneficial—especially if we are to understand how class and racialization tug at the mechanics of language contact.

7.2 Rhythm in Södersnack

While the three-way split for rhythm is interesting from a socio-systematic perspective, it is also interesting in light of the paucity of studies on Södersnack. Section 2.2 offered a review of Stockholm's industrial-era working-class variety Södersnack and proposed that it might have higher rhythmic alternation than the city's other varieties due to its higher inventory of diphthongal long vowels (Kotsinas, 1994; Öqvist, 2010). A similar link has been proposed by Torgersen and

Szakay (2012), albeit in the other direction. The typically diphthongal FACE and GOAT vowels are known to be monophthongal in MLE, so a link was suggested between these and the low-nPVIV findings for MLE.

Important here also is the result that Södersnack is considerably *more* rhythmically deviant from received Stockholmian than multiethnolect. Whereas numerous accounts of staccato rhythm circulate for Stockholm's multiethnolect, I am unaware of any such metalinguistic account for the prosody of Södersnack. Furthermore, there has been an underlying assumption in any discourse on Swedish multiethnolect that the distance between its variants and standard variants would surely be larger than the distance between more "indigenous" varieties and the standard. The present findings disrupt this assumption and typologically place Stockholm's multiethnolect closer to the received standard than Södersnack as far as rhythm is concerned.

The implication of the the strong tendency to style-shift among the *svensk* working class also implicates a degree of salience on high-alternation rhythm. But again, the lack of scholarly and popular commentary about its prosody frustrates any meaningful interpretation. So rather than rhythm being a sociolinguistic *marker* in the epiphenomenal sense, a more likely possibility is that it is a phonotactic result of Södersnack diphthongal segments being replaced with received-standard monophthongal counterparts in more formal speech. This link has neither been tested nor substantiated, but the proposal is not unreasonable given what is known about its vowels (Bergman, 1946; Kotsinas, 1994; Öqvist, 2010; Ståhle, 1975).

7.3 Social salience

Turning to the question of social salience, the respective rhythmic variants of both working-class groups appear to target the rhythmic pattern of higher-class casual speech. The focusing is less uniform in the unsolicited reading task (READING/*formal*) and more focused in the solicited reading task (RADIO/*very formal*). For *all svensk*—along with the upper-middle class *invandrare* speakers—a shared norm seems to already be the target in the unsolicited reading task. For *invandrare* speakers lower in the class hierarchy, this does not seem to be the case. Rather, the additional prompt of "sound like an announcer on Radio Sweden" seemed to be necessary for that additional nudge toward the community's hegemonic norm. This is also something that is audible in many speech samples. An example is speaker Hayder (Figure 2, bottom left corner), whose unprompted READING style sounds reasonably similar to how he speaks with his peers. His RADIO style, however, sounds like an entirely different register (when I played it back for his friends, they were in disbelief). This is not to say that Hayder would ever actually *use* that register, but it reveals something about a latent ability to release certain features and move closer to features found higher in the social-class hierarchy.

As discussed above, the variant that belongs to the *svensk* working class appears to be much more socially salient than the variant that was of primary interest to this study, namely the staccato variant of the *invandrare* working class. If one accepts the *indicator > marker > stereotype* progression proposed by Labov (1972), this would imply that the former is a newer feature than the latter, something that bolsters the possibility that the former is a legacy feature of the city's industrial working class.

As it concerns the social salience of staccato rhythm, the mild degree of style-shifting suggests that the feature still has mild social salience for its speakers. If the simulation in Figure 4 is re-examined, lower-middle class *invandrare* speakers are predicted to style-shift into the intermediate rhythmic pattern produced by higher social groups. The lowest *invandrare* class is predicted by the model to shift by the same amount, but their lower start point means that full normative rhythm is not achieved. If one considers these findings in light of those of Bijvoet and Fraurud (2011; reviewed in Section 2.5), one could imagine that the mild shift predicted for a lower working-class speaker would be sufficiently prestige-sounding for his peers while remaining deviant-sounding for higher-class speakers. So if one looks at such a shift from 46.8 to 48.8 for a working-class *invandrare* born in 1997 (Figure 4), this may register as "good Swedish" for a number of multiethnolectal speakers while still sounding like "Rinkeby Swedish" for speakers from other social groups.

Returning to the evolutionary progression proposed by Labov (1972), Table 5 Model 2 offers appealing results because the added interaction of style with age renders an increase in stylistic sensitivity as the age of *invandrare* speakers decreases. This implies a continued movement from indicator to marker. However, as discussed in Section 5, the age and style interaction had to ultimately be discarded because it weakened the model fit and was not significant. The direction of the coefficients, however, serves as a reminder to test stylistic sensitivity in apparent time in future investigations.

7.4 Rhythm in apparent time

When examining age as an apparent-time proxy for diachronic development (Labov, 2001), the statistical model reveals two important trends: (a) the staccato low-alternation feature of working-class *invandrare* men is becoming more staccato over time, and (b) the speech of *all* male groups is becoming more staccato over time while maintaining similar stratifications, illustrated most clearly in Figure 4. To rephrase the last point, the whole of Stockholm appears to be becoming staccato over time, led by the *invandrare* working-class, while the *svensk* working class appears to lag the furthest behind. A higher-class speaker born in 1987 (53.8 to 56.7) is predicted to have speech rhythm that resembles a working-class and lower-middle-class *invandrare* born in 1977 (54.6 and 56.4, respectively). A higher-class speaker born in 1997 (49.9 to 52.8) is predicted to speak with a rhythm that is actually lower than that of a working-class *invandrare* speaker born in 1977 (54.6).

Two possible reasons come to mind for this trend. The first possibility is that the contact prosody of Stockholm's lower classes is an active change from below that is incrementally diffusing into mainstream speech all while younger speakers "hypercorrect from below" (Labov, 1972, p. 178) by moving even further into staccato territory. A similar process was found by Trudgill (1988) for t-glottaling in Norwich. T-glottaling was an exogenous contact feature from the South that first entered Norwich through the working-class vernacular. With time, it climbed the class hierarchy and became more or less socially ubiquitous (p. 45).

A second possibility is that extensive language contact, beyond that of the multiethnic periphery, is rendering Swedish prosody less typologically marked. Sweden is characterized by its many global brands and its robust export economy, and Stockholm lies at the center of this activity. Further, the city has witnessed a recent finance and technology boom that has brought in expatriates from all corners of the globe. When comparing nPVIV of duration for Swedish (54.9, Young, 2018b, p. 50) against the other languages that have been studied (Fuchs, 2014a, p. 81), only British, New Zealand, and Thai English have higher nPVIV values. All other languages tested in the literature have lower nPVIV measurements. Since English and Swedish are the two dominant lingua francas in Stockholm, it is plausible that contact with L1-accented Swedish, L1-accented English, and American English may be driving this incremental downward shift in the rhythmic alternation of Stockholmian prosody.

7.5 Accounting for rhythm phonotactically

As it pertains to phonotactic accounts, these findings lead to interesting questions. Table 7 demonstrated that the internally-governed variation within a single language like Swedish is extremely high for nPVIV. When an accent-1 long vowel is adjacent to an unstressed short vowel, the nPVIV values are predicted to be very high; when an unstressed long vowel is adjacent to an unstressed long vowel, the nPVIV values are predicted to be very low. Although the sociolinguistic analysis has revealed socially-governed variation that is independent of internal factors, it has not investigated where in the phonology the change is occurring. As discussed in Section 2.4, Kotsinas speculated that Rinkeby Swedish might be reducing the difference between long and short syllables (Kotsinas, 1988a, p. 268; Kotsinas, 1990, p. 257) and that my preliminary investigation shows that this may be the case (Young, 2019, p. 213). Are working-class *invandrare* men reducing their accented vowels durationally? Or are they enacting a reduction in, say, intensity? Or is it rather the case that the staccato effect is due to the enlargement—be it in duration, F0, intensity, or all three combined—of unstressed vowels? Or is it the case that the rhythmic grid is epiphenomenal to its segmental components, enacting enlargements or reductions wherever necessary to coerce a specific pattern? These questions are currently under investigation, and the results are still tentative (Young, 2019, pp. 189–198).

As was reviewed in Section 2.4, a number of other options are possible such as monophthongal long vowels (Young, 2019, p. 209), unstressed vowels qualitatively further from schwa (Young, 2019, p. 212), phrase-final lengthening (Young, 2019, pp. 197–198), and elided or flapped coda rhotics (Young, 2018b, pp. 50–51) among the *invandrare* working class. None of these findings have gone through the peer-review process and are still preliminary in nature. Furthermore, researching and substantiating their presence would only be the first step. The second step would be to test their correlation with the current stratification of rhythmic contrast. The present study, however, has established rhythm as a stratified variable in Stockholm and can hopefully be a point of departure for future studies that might wish to explore related phonotactic phenomena.

Acknowledgements

I wish to thank Devyani Sharma, Erez Levon, Michael McGarrah, Paul Kerswill, Nicolai Pharao, Adam Chong, and Erik Thomas for the helpful and generous feedback that led to this final analysis. Thanks also to Keder Akman and Laura Yu for helping recruit participants and to the participants themselves for being so generous with their time. I am also grateful to the anonymous reviewers who dedicated so much time and effort to ensure that this work might meet the rigor of our discipline. I am of course responsible for all remaining shortcomings.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was funded by The British Economic and Social Research Council and the Sven och Dagmar Salén Foundation.

ORCID iD

Nathan J. Young https://orcid.org/0000-0002-0337-568X

Notes

1. The term is troubled in Section 2.1.
2. One such feature, for example, is the “damped” /iː/ discussed in Gross (2018, pp. 319–320).
3. To offer a case in point, one of the upper-class participants in this study, Johan, reported in his interview that when he played Monopoly with his grandmother as a child, she refused to buy property on any Södermalm streets—even if it meant forfeiting the game.
4. In Stockholm, “förort” (suburb) has the same connotation as the North American term “urban” or “inner-city” because the racialized working classes live in the city’s suburban periphery.
5. According to Education First, Sweden ranks number 2 (after the Netherlands) for English proficiency out of 100 nations that do not have English as an official language.

6. This is not always the case in the data. Trills and taps do occasionally occur, which further necessitates accounting for any R-ful observation in the statistical model.
7. Low et al. (2000) called it *pairwise variability index of vowels* (PVI) because the normalization and vocalic aspects of the algorithm were presupposed. It was only later, after non-normalized and non-vocalic spinoffs emerged, that “normalized” and “vowel” were added to the title.
8. For example a 60 ms vowel followed by a 120 ms vowel produces an nPVI of $\frac{|(120-60)|}{90} = 0.67$ and an RIM of $|ln\frac{60}{120}| = 0.69$. A 30 ms vowel followed by a 120 ms vowel produces an nPVI of $\frac{|(120-30)|}{75} = 1.20$ and an RIM of $|ln\frac{30}{120}| = 1.38$. The two algorithms only render widely-different outputs in outlier situations (like a 10 ms vowel followed by a 120 ms vowel, in which case the RIM will skew and render a much higher contrast index than the nPVI).
9. It is my view that how one measures rate is less important when analyzing a single language. Syllable duration, syllables per second, segmental duration, or segments per second all seem to be adequate approaches. The choice of measurement gains importance only *once one cross-compares languages*, because they can have different syllable structures (consider languages that have CV syllables versus CCCVCCC). See Dellwo (2010, p. 67) for a similar discussion.
10. The symbol : is the annotation in *R* for a statistical interaction.
11. This does not mean that I am ruling out the possibility that age interacts with class and race in Stockholm; rather I am concluding that the current dataset cannot reliably say anything about that interaction because it lacks sufficient orthogonality.

References

Aktürk-Drake, M. (2018). Hur bra har den *svenska* integrationskontexten varit på att främja balanserad tvåspråkighet? [How well has the Swedish integration context been for advancing balanced bilingualism?]. *Nordand – Nordic Journal of Bilingualism Research*, *2*(02), 107–130.

Allwood, J. (1999). The Swedish spoken language corpus at Göteborg University. In *Proceedings Fonetik 99: The Swedish Phonetics Conference*, June 1999. *Gothenburg University*.

Arvaniti, A. (2009). Rhythm, timing and the timing of rhythm. *Phonetica*, *66*(1–2), 46–63.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

Bauman, Z. (2000). *Liquid modernity*. Polity Press.

Bergman, G. (1946). Språket på Söder [The language on Södermalm]. In G. Bergman (Ed.), *Ord och Stil* [Words and style] (pp. 100–119). Prisma.

Bijvoet, E., & Fraurud, K. (2008). Svenskan i dagens flerspråkiga storstadsmiljöer: En explorativ studie av unga stockholmares perceptioner av variation och varieteter [Swedish in today’s multilingual enviroments: An explorative study of young Stockholmers’ perceptions of variation and varieties]. *Nordand – Nordic Journal of Bilingualism Research*, *3*(2), 7–38.

Bijvoet, E., & Fraurud, K. (2011). Language variation and varieties in contemporary multilingual Stockholm: An exploratory pilot study of young people’s perceptions. In R. Källström & I. Lindberg (Eds.), *Young urban Swedish: Variation and change in multilingual settings. Göteborgsstudier i nordisk språkvetenskap, 14* (pp. 1–34). Gothenburg University Press.

Bijvoet, E., & Fraurud, K. (2012). Studying high-level (L1–L2) development and use among young people in multilingual Stockholm. *Studies in Second Language Acquisition*, *34*(2), 291–319.

Bijvoet, E., & Fraurud, K. (2016). What’s the target? A folk linguistic study of young Stockholmers’ constructions of linguistic norm and variation. *Language Awareness*, *25*(1–2), 17–39.

Bodén, P. (2007). ‘Rosengårds*svensk*’ fonetik och fonologi [‘Rosengård Swedish’ phonetics and phonology]. In L. Ekberg (Ed.), *Språket hos ungdomar i en flerspråkig miljö i Malmö - Nordlund 27* [Youth language in a multilingual milieu in Malmö] (pp. 1–47). Lund University Press.

Boersma, P., & Weenink, D. (2017). *Praat: doing phonetics by computer* [*Computer software*], *Version 6.0.36*. http://www.praat.org/

Brato, T. (2015, July 14). *TB-Basic Vowel Analysis, Version 2.2* [*Computer software*]. http://www.gnu.org/licenses/gpl.txt

Breen, M., Fedorenko, E., Wagner, M., & Gibson, E. (2010). Acoustic correlates of information structure. *Language and Cognitive Processes*, *25*(7–9), 1044–1098.

Cerquiglini, B. (2001). Le français d'aujourd'hui, ça bouge [Today's French is on the move]. *Construire*, 7.

Cheshire, J. (2013). Grammaticalisation in social context: The emergence of a new English pronoun. *Journal of Sociolinguistics*, *17*(5), 608–633.

Clopper, C. G., & Smiljanic, R. (2015). Regional variation in temporal organization in American English. *Journal of Phonetics*, *49*, 1–15.

Clyne, M. (2000). Lingua franca and ethnolects in Europe and beyond. *Sociolinguistica*, *14*, 83–89.

Coggshall, E. L. (2008). The prosodic rhythm of two varieties of Native American English. *University of Pennsylvania Working Papers in Linguistics*, *14*(2), 1–10.

Cornips, L., & de Rooij, V. A. (2013). Selfing and othering through categories of race, place, and language among minority youths in Rotterdam, The Netherlands. In S. Peter, I. Gogolin, M. E. Schulz, & J. Davydova (Eds.), *Multilingualism and language diversity in urban areas: Acquisition, identities, space, education* 1 (pp. 129–164). John Benjamins.

Cumming, R. E. (2010). *Speech rhythm: The language-specific integration of pitch and duration* [Doctoral dissertation, University of Cambridge]. https://www.repository.cam.ac.uk/handle/1810/228685

Dellwo, V. (2006). Rhythm and speech rate: A variation coefficient for deltaC. *Language and Language Processing: Proceedings of the 38th Linguistic Colloquium, Peter Lang*, 231–241.

Dellwo, V. (2010). *Influences of speech rate on the acoustic correlates of speech rhythm: An experimental phonetic study based on acoustic and perceptual evidence* [Doctoral dissertation, Rheinischen Friedrich-Wilhelms-Universität]. https://www.researchgate.net/profile/Volker_Dellwo/publication/280239421_Influences_of_speech_rate_on_the_acoustic_correlates_of_speech_rhythm_An_experimental_phonetic_study_based_on_acoustic_and_perceptual_evidence/links/55aeeb0108aee0799220ea29/Influences-of-speech-rate-on-the-acoustic-correlates-of-speech-rhythm-An-experimental-phonetic-study-based-on-acoustic-and-perceptual-evidence.pdf

Deterding, D. (2001). The measurement of rhythm: A comparison of Singapore and British English. *Journal of Phonetics*, *29*(2), 217–230.

Doran, M. (2001). Negotiating between Bourge and Racaille: Verlan as youth identity practice in suburban Paris. In A. Pavlenko & A. Blackledge (Eds.), *Negotiation of identities in multilingual contexts* (pp. 93–124). Multilingual Matters.

Duez, D., & Casanova, M.-H. (1997). Quelques aspects de l'organisation temporelle du parlé des banlieues parisiennes [Some aspects of the temporal organization of speech in the Paris suburbs]. *Revue Parole*, *1*, 59–74.

Engstrand, O. (1990). Swedish. *Journal of the International Phonetic Association*, *20*(1), 42–44.

Engstrand, O., Bruce, G., Elert, C.-C., Eriksson, A., & Strangert, E. (2000). *Databearbetning i SweDia 2000: segmentering, transkription och taggning. Version 2.2* [Data work in SweDia 2000: transcription, segmentation, and tagging]. University of Gothenburg. https://docplayer.se/47215375-Databearbetning-i-swedia-2000-segmentering-transkription-och-taggning-version-2-2.html

Experian Ltd, & InsightOne Nordic AB. (2013). *Mosaic Sweden e-handbook: The classification of Swedish consumers*. Stockholm: InsightOne AB.

Fagyal, Z. (2010). Rhythm types and the speech of working-class youth in a banlieue of Paris: The role of vowel elision and devoicing. In D. R. Preston & N. Niedzielski (Eds.), *A reader in sociophonetics* (pp. 91–132). Mouton de Gruyter.

Fant, G., & Kruckenberg, A. (1994). Notes on stress and word accent in Swedish. *Department for Speech, Music and Hearing, Quarterly Progress and Status Report*, 125–144.

Fant, G., Kruckenberg, A., & Liljencrants, J. (2000). Acoustic-phonetic analysis of prominence in Swedish. In A. Botinis (Ed.), *Intonation: Analysis, modelling and technology* (pp. 55–86). Springer.

Fontanella de Weinberg, M. B. (1974). *Un aspecto sociolingüístico del español bonaerense: La -s en Bahía Blanca* [A sociolinguistic feature in Bonaerense Spanish: The -s in Bahia Blanca]. Universidad Nacional del Sur.

Forsberg, H. (2018). School competition and social stratification in the deregulated upper secondary school market in Stockholm. *British Journal of Sociology of Education*, *39*(6), 891–907.

Fraisse, P. (1963). *The psychology of time*. Harper & Row.

Fraisse, P. (1982). Rhythm and tempo. In D. Deutsch (Ed.), *The psychology of music* (pp. 149–180). Academic Press.

Fraurud, K. (2003). Svenskan i Rinkeby och andra flerspråkiga bostadsområden [Swedish in Rinkeby and other multilingual residential environments]. *Sprog i Norden*, *34*(1), 77–92.

Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. *The Journal of the Acoustical Society of America*, *27*(4), 765–768.

Fry, D. B. (1958). Experiments in the perception of stress. *Language and Speech*, *1*(2), 126–152.

Fuchs, R. (2014a). Integrating variability in loudness and duration in a multidimensional model of speech rhythm: Evidence from Indian English and British English. In *Proceedings of Speech Prosody 2014* (pp. 290–294).

Fuchs, R. (2014b). Towards a perceptual model of speech rhythm: Integrating the influence of f0 on perceived duration. In *Proceedings of INTERSPEECH 2014* (pp. 1949–1953).

Fuchs, R. (2014c). You got the beat: Rhythm and timing. In R. Monroy-Casas & I. Arboleda-Guirao (Eds.), *Readings in English phonetics and phonology* (pp. 165–188). IULMA Valencia.

Fuchs, R. (2016). *Speech rhythm in varieties of English: Evidence from educated Indian English and British English*. Springer.

Gabriel, C., & Kireva, E. (2014). Prosodic transfer in learner and contact varieties: Speech rhythm and intonation of Buenos Aires Spanish and L2 Castilian Spanish produced by Italian native speakers. *Studies in Second Language Acquisition*, *36*(2), 257–281.

Galves, A., Garcia, J., Duarte, D., & Galves, C. (2002). Sonority as a basis for rhythmic class discrimination. In *Proceedings of Speech Prosody 2002* (pp. 323–326).

Ganzeboom, H. B., & Treiman, D. J. (2003). Three internationally standardised measures for comparative research on occupational status. In J. H. Hoffmeyer-Zlotnik & C. Wolf (Eds.), *Advances in cross-national comparison: A European working book for demographic and socio-economic variables* (pp. 159–193). Kluwer Academic/Plenum Publishers.

Ganzeboom, H. B., & Treiman, D. J. (2018). *International stratification and mobility file: Conversion tools*. Department of Social Research. http://www.harryganzeboom.nl/ismf/index.htm

Gates, S. M. (2018). Why the long FACE?: Ethnic stratification and variation in the London diphthong system. *University of Pennsylvania Working Papers in Linguistics*, 24(2), 38–48. https://repository.upenn.edu/pwpl/vol24/iss2/6/

Gibbon, D. (2003). Computational modelling of rhythm as alternation, iteration and hierarchy. In M.-J. Solé & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences, 3–9 August 2003, Barcelona, Spain* (pp. 2489–2492). Universitat Auto'noma de Barcelona.

Gibbon, D., & Gut, U. (2001). Measuring speech rhythm. In P. Dalsgaard, B. Lindberg, H. Benner, & Z.-h. Tan (Eds.), *Proceedings of Eurospeech 2001* (pp. 91–94). Center for Personkommunikation.

Gilbers, S., Hoeksema, N., de Bot, K., & Lowie, W. (2019). Regional variation in West and East Coast African-American English prosody and rap flows. *Language and Speech*. doi:10.1177/0023830919881479

Grabe, E, & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. *Papers in Laboratory Phonology*, *7*, 515–546.

Gross, J. (2018). Segregated vowels: Language variation and dialect features among Gothenburg youth. *Language Variation and Change*, *30*(3), 315.

Hall, T., & Vidén, S. (2005). The million homes program: A review of the great Swedish planning project. *Planning Perspectives*, *20*(3), 301–328.

Hansen, G. F., & Pharao, N. (2010). Prosody in the Copenhagen multiethnolect. In P. Quist & B. A. Svendsen (Eds.), *Multilingual urban Scandinavia: New linguistic practices* (pp. 79–95). Multilingual Matters.

He, L. (2012). Syllabic intensity variations as quantification of speech rhythm: Evidence from both L1 and L2. In Q. Ma, H. Ding, & D. Hirst (Eds.), *Proceedings of Speech Prosody 2012, Sixth International Conference* (pp. 466–469). Tongji University Press.

Heldner, M. (2003). On the reliability of overall intensity and spectral emphasis as acoustic correlates of focal accents in Swedish. *Journal of Phonetics*, *31*(1), 39–62.

Heldner, M., & Strangert, E. (1997). To what extent is perceived focus determined by F0-cues? In G. Kokkinakis, F. Nikos, & E. Dermatas (Eds.), *Proceedings of the Fifth European Conference on Speech Communication and Technology, EUROSPEECH 1997* (pp. 875–877). ESCA.

Heo, M., & Leon, A. C. (2010). Sample sizes required to detect two-way and three-way interactions involving slope differences in mixed-effects linear models. *Journal of Biopharmaceutical Statistics*, *20*(4), 787–802.

Hesse, B. (2007). Racialized modernity: An analytics of white mythologies. *Ethnic and Racial Studies*, *30*(4), 643–663.

Holmér, J. (2014). *Subway stations small.jpg*. Wikipedia Commons. https://sv.rn.wikipedia.org/wiki/Fil:Subway_stations_small.jpg

Hübinette, T., Hörnfeldt, H., Farahani, F., & Rosales, R. L. (2012). Om ras och vithet i det samtida Sverige [On race and whiteness in contemporary Sweden]. In T. Hübinette, H. Hörnfeldt, F. Farahani, & R. L. Rosales (Eds.), *Om ras och vithet i det samtida Sverige* [On race and whiteness in contemporary Sweden] (pp. 11–36). Mångkulturellt Centrum.

Johnson, D. E. (2014). *Progress in regression: Why natural language data calls for mixed-effects models*. http://www.danielezrajohnson.com/johnson_2014b.pdf

Keim, I. (2007). Socio-cultural identity, communicative style, and their change over time: A case study of a group of German-Turkish girls in Mannheim/Germany. In P. Auer (Ed.), *Style and social identities: Alternative approaches to linguistic heterogeneity* (vol. 18) (pp. 155–186). Mouton de Gruyter.

Keim, I., & Knöbl, R. (2007). Sprachliche Varianz und sprachliche Virtuosität türkisch-stämmiger Ghetto-Jugendlicher in Mannheim [Linguistic variation and linguistic virtuosity of Turkish-heritage ghetto youth in Mannheim]. In C. Fandrych & R. Salverda (Eds.), *Standard, variation und Sprachwandel in germanischen Sprachen* [Standard, variation and language change in Germanic languages] (pp. 157–199). Gunther Narr.

Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *The Journal of the Acoustical Society of America*, *59*(5), 1208–1221.

Kochanski, G., Grabe, E., Coleman, J., & Rosner, B. (2005). Loudness predicts prominence: Fundamental frequency lends little. *The Journal of the Acoustical Society of America*, *118*(2), 1038–1054.

Kotsinas, U.-B. (1988a). Rinkebysvenska - en dialekt? [Rinkeby Swedish – a dialect?]. In P. Linell, V. Adelswärd, T. Nilsson, & P. A. Pettersson (Eds.), *Svenskans beskrivning 16* [The description of Swedish 16] (vol. 1, pp. 264–278). Tema Kommunikation. http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-108104

Kotsinas, U.-B. (1988b). Stockholmspråk i förändring [Stockholm's language in change]. In G. Pettersson (Ed.), *Studier i svensk språkhistoria* [Studies in Swedish language history] (vol. 1, pp. 133–147). Lund University Press.

Kotsinas, U.-B. (1990). Svensk, Invandrarsvensk eller Invandrare? Om Bedömning av "Främmande" Drag i Ung-domsspråk [Swedish, immigrant Swedish or immigrant? On assessments of "foreign" features in youth language]. In G. Tingbjörn (Ed.), *Andra symposiet om svenska som andraspråk i Göteborg 1989* [Second Symposium on Swedish as a Second Language in Gothenburg 1989] (pp. 244–275). Scriptor.

Kotsinas, U.-B. (1994). Snobbar och pyjamastyper: Ungdomskultur, ungdomsspråk och gruppidentiteter i Stock-holm [Snobs and pyjama types: Youth culture, youth language and group identities in Stockholm]. In J. Fornäs, U. Boethius, M. Forsman, H. Ganetz, & B. Reimer (Eds.), *Ungdomskultur i Sverige* [Youth culture in Sweden] (pp. 311–336). Brutus Östlings Bokförlag.

Labov, W. (1963). The social motivation of a sound change. *Word*, *19*(3), 273–309.

Labov, W. (1972). *Sociolinguistic patterns*. University of Pennsylvania Press.

Labov, W. (2001). *Principles of linguistic change, volume 2: Social factors*. Blackwell.

Labov, W. (2006 [1966]). *The social stratification of English in New York City* (2nd ed.). Cambridge University Press.

Lee, R. L. (2010). On the margins of belonging: Confronting cosmopolitanism in the late modern age. *Journal of Sociology*, *46*(2), 169–186.

Lehtonen, H. (2011). Developing multiethnic youth language in Helsinki. In F. Kern & M. Selting (Eds.), *Ethnic styles of speaking in European metropolitan areas* (pp. 291–318). John Benjamins.

Lenneberg, E. H. (1967). *The biological foundations of language*. Wiley and Son.

Lentin, A. (2008). Europe and the silence about race. *European Journal of Social Theory*, *11*(4), 487–503.

Lentin, A., & Titley, G. (2011). *The crises of multiculturalism: Racism in a neoliberal age*. Zed Books.

Lerdahl, E., & Jackendoff, R. A. (1983). *Generative theory of tonal music*. MIT Press.

Lieberman, P. (1960). Some acoustic correlates of word stress in American English. *The Journal of the Acoustical Society of America*, *32*(4), 451–454.

Low, E. L. (1998). *Prosodic prominence in Singapore English* [Doctoral dissertation, University of Cambridge]. https://www.repository.cam.ac.uk/handle/1810/251470

Low, L. E., Grabe, E., & Nolan, F. (2000). Quantitative characterizations of speech rhythm: Syllable-timing in Singapore English. *Language and Speech*, *43*(4), 377–401.

Mendoza-Denton, N. (2008). *Homegirls: Language and cultural practice among Latina youth gangs*. Wiley-Blackwell.

Milani, T. M., & Jonsson, R. (2012). Who's afraid of Rinkeby Swedish? Stylization, complicity, resistance. *Journal of Linguistic Anthropology*, *22*(1), 44–63.

Mishra, T., Sridhar, V. R., & Conkie, A. (2012). Word prominence detection using robust yet simple prosodic features. In *Proceedings of the 13th Annual Conference of the International Speech Communication Association 2012 (INTERSPEECH 2012)* (pp. 1–4). ISCA.

Morris, U., & Zetterman, H. (2011). *Från bondgård till cirkus. Konstruktion av en högläsningstext för bedömning av röst-och talfunktion och talandning* [From farm to circus: Constructing a reading passage for assessment of voice and speech faculties] [Master's thesis, Department of Speech Therapy, Karolinska Institutet]. https://clintec.ki.se/Exam_logopedi/pdf/338.pdf

Mulinari, D., & Neergaard, A. (2004). *Den nya svenska arbetarklassen – Facket och de rasifierade arbetarna*. [The new Swedish working class: Trade unions and racialized workers]. Borea.

Nordenstam, K., & Wallin, I. (2002). *Osynliga flickor – synliga pojkar: Om ungdomar med svenska som andraspråk* [Invisible girls – visible boys: On youth with Swedish as a second language]. Studentlitteratur.

Öqvist, J. (2010). Riktig stockholmska? Ekenssnacket som varietet och fenomen [Real Stockholmian? Ekenssnack as a variety and phenomenon]. In M. Reinhammar, L. Elmevik, S. Fridell, M. Thelander, & H. Williams (Eds.), *Studier i svenska språkets historia* [Studies in the history of the Swedish Language], *11 (Acta Academiae Regiae Gustavi Adolphi CXIII)* (pp. 253–260). Kungl. Gustav Adolfs Akademien för svensk folkkultur.

Payne, E., Post, B., Astruc, L., Prieto, P., & Vanrell, M. d. M. (2012). Measuring child rhythm. *Language and Speech*, *55*(2), 203–229.

Pharao, N., & Maegaard, M. (2017). On the influence of coronal sibilants and stops on the perception of social meanings in Copenhagen Danish. *Linguistics*, *55*(5), 1141–1167.

Pharao, N., Maegaard, M., Møller, J. S., & Kristiansen, T. (2014). Indexical meanings of [s+] among Copenhagen youth: Social perception of a phonetic variant in different prosodic contexts. *Language in Society*, *43*(1), 1–31.

Pluymaekers, M., Ernestus, M., & Baayen, R. H. (2005). Lexical frequency and acoustic reduction in spoken Dutch. *The Journal of the Acoustical Society of America*, *118*(4), 2561–2569.

Prieto, P., del Mar Vanrell, M., Astruc, L., Payne, E., & Post, B. (2012). Phonotactic and phrasal properties of speech rhythm. Evidence from Catalan, English, and Spanish. *Speech Communication*, *54*(6), 681–702.

Quist, P. (2008). Sociolinguistic approaches to multiethnolect: Language variety and stylistic practice. *International Journal of Bilingualism*, *12*(1–2), 43–61.

Quist, P. (2010). Untying the language, body and place connection: Linguistic variation and social style in a Copenhagen community of practice. In P. Auer & J. E. Schmidt (Eds.), *Handbücher Zur Sprach- Und Kommunikationswissenschaft* [An international handbook of linguistic variation. Vol. 1: Theories and methods] (pp. 632–648). Mouton de Gruyter.

Quist, P. (2012). Skandinavisk i kontakt med indvandrersprog [Scandinavian in contact with immigrant languages]. *Sprog i Norden*, *43*(1), 1–14. https://tidsskrift.dk/sin/article/download/17154/14896

Rampton, B. (2006). *Language in late modernity: Interaction in an urban school*. Cambridge University Press.

Rampton, B. (2011). From 'multi-ethnic adolescent heteroglossia' to 'contemporary urban vernaculars'. *Language & Communication*, *31*(4), 276–294.

Ramus, F., Nespor, F., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, *73*, 265–292.

Riad, T. (2014). *The phonology of Swedish*. Oxford University Press.

Sarmah, P., Gogoi, D. V., & Wiltshire, C. (2011). Thai English – Rhythm and vowels. In L. Lim & N. Gisborne (Eds.), *Typology of Asian Englishes* (pp. 75–96). Benjamins.

Scott, D. R., Isard, S., & de Boysson-Bardies, B. (1986). On the measurement of rhythmic irregularity: A reply to Benguerel. *Journal of Phonetics*, *14*(2), 327–330. https://doi.org/10.1016/S0095-4470(19)30659-X

Sharma, D., & Rampton, B. (2015). Lectal focusing in interaction: A new methodology for the study of style variation. *Journal of English Linguistics*, *43*(1), 3–35.

Sharma, D., & Sankaran, L. (2011). Cognitive and social forces in dialect shift: Gradual change in London Asian speech. *Language Variation and Change*, *23*(3), 399–428.

Shousterman, C. (2015). *Speaking English in Spanish Harlem: Language change in Puerto Rican English* [Doctoral dissertation, New York University]. https://search.proquest.com/docview/1666862501?pq-origsite=gscholar&fromopenview=true

Ståhle, C. I. (1975). En undersökning av Stockholmsspråk [An investigation of Stockholm speech]. In C. Ståhle (Ed.), *Stockholmsnamn och Stockholmsspråk* [Stockholm names and Stockholm speech] (pp. 81–89). Norstedts.

Strangert, E., & Heldner, M. (1995). The labelling of prominence in Swedish by phonetically experienced transcribers. In K. Elenius & P. Branderud (Eds.), *Proceedings of the Ninth International Congress of Phonetic Sciences (ICPhS)* (vol. 4, pp. 204–207). Royal Institute of Technology and Stockholm University.

Sveriges Television. (1979, October 24). *Invandrarungdomar: En tidsinställd bomb* [Immigrant youth: A ticking time-bomb] [Television news report]. P1 Studio S in TV1 Lektionskväll.

Tagliamonte, S. A. (2006). *Analysing sociolinguistic variation*. Cambridge University Press.

Therborn, G. (1998). A unique chapter in the history of democracy: The Swedish social democrats. In K. Misgeld, K. Molin, & K. Åmark (Eds.), *Creating social democracy* (pp. 1–34). Penn State University Press.

Thesleff, A. (1912). *Stockholms forbrytarspråk och lägre slang, 1910–1912* [Stockholm's criminal language and lower slang, 1910–1912]. Albert Bonniers Förlag.

Thomas, E., & Carter, P. (2006). Prosodic rhythm and African American English. *English World-Wide*, *27*, 331–355.

Topografiska Corpsen. (1861). Trakten omkring Stockholm i iX blad [The area around Stockholm in 9 pages]. Stockholm: Stockholms Stadsarkiv. https://stockholmskallan.stockholm.se/post/31498

Torgersen, E., & Szakay, A. (2012). An investigation of speech rhythm in London English. *Lingua*, *122*(7), 822–840.

Trudgill, P. (1974). *The social differentiation of English in Norwich*. Cambridge University Press.

Trudgill, P. (1988). Norwich revisited: Recent linguistic changes in an English urban dialect. *English World-Wide*, *9*(1), 33–49.

Turk, A. E., & Sawusch, J. R. (1996). The processing of duration and intensity cues to prominence. *The Journal of the Acoustical Society of America*, *99*(6), 3782–3790.

Wacquant, L. (2008). *Urban outcasts: A comparative sociology of advanced marginality*. Polity.

Wagner, M., & Watson, D. G. (2010). Experimental and theoretical advances in prosody: A review. *Language and Cognitive Processes*, *25*(7–9), 905–945.

Watt, P., Millington, G., & Huq, R. (2014). East London mobilities: The 'Cockney Diaspora' and the remaking of the Essex ethnoscape. In P. Watt & P. Smets (Eds.), *Mobilities and neighbourhood belonging in cities and suburbs* (pp. 121–144). Springer.

White, L., & Mattys, S. L. (2007). Calibrating rhythm: First language and second language studies. *Journal of Phonetics*, *35*(4), 501–522.

White, L., Mattys, S. L., & Wiget, L. (2012). Language categorization by adults is based on sensitivity to durational cues, not rhythm class. *Journal of Memory and Language*, *66*(4), 665–679.

Wiese, H. (2012). Kiezdeutsch: Ein neuer Dialekt entsteht [Hood German: A new dialect emerges]. C.H. Beck.

Wiget, L., White, L., Schuppler, B., Grenon, I., Rauch, O., & Mattys, S. L. (2010). How stable are acoustic metrics of contrastive speech rhythm? *The Journal of the Acoustical Society of America*, *127*(3), 1559–1569.

Wolfram, W. (1969). *A sociolinguistic description of Detroit Negro speech, urban language series, no. 5*. Center for Applied Linguistics.

Woodrow, H. (1951). Time perception. In S. S. Stevens (Ed.), *Handbook of experimental psychology* (pp. 1224–1236). Wiley.

Young, N. (2018a). 'Copycats, ja dom shouf': Using hip hop to compare lexical replications in Danish and Swedish multiethnolects. *University of Pennsylvania Working Papers in Linguistics*, 24(2), 174–184.

Young, N. (2018b). Talrytmens sociala betydelse i det senmoderna Stockholm [The social meaning of speech rhythm in late-modern Stockholm]. *Nordand – Nordic Journal of Bilingualism Research*, *13*(1), 41–63. https://docs.wixstatic.com/ugd/5215a0_b6223326a63645c4b35811dc4b1501c4.pdf

Young, N. (2019). *Rhythm in late-modern Stockholm: Social stratification and stylistic variation in the speech of men* [Doctoral dissertation, Department of Linguistics, Queen Mary, University of London]. http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-178897

Young, N. (in press, 2021). 'Benim' – A new pronoun in Swedish. In H. Van de Velde, N. Haug Hilton, & Knooihuizen (Eds.), *Studies in language variation: Selected papers of ICLaVE10*. John Benjamins.

Young, N., & McGarrah, M. (2017, November 2–5). Introducing NordfA: Forced alignment of Nordic languages. *Presentation at New Ways of Analyzing Variation (NWAV) 46*, *Madison, USA*.

Zhao, Y., & Jurafsky, D. (2009). The effect of lexical frequency and Lombard reflex on tone hyperarticulation. *Journal of Phonetics*, *37*(2), 231–247.