The acoustic characteristics of Swedish vowels

Anna Persson[1]

[1] Department of Swedish Language and Multilingualism, Stockholm University

Author Note

Correspondence concerning this article should be addressed to Anna Persson, Department of Swedish Language and Multilingualism, Stockholm University. E-mail: anna.persson@su.se

Abstract

The Swedish vowel space is relatively densely populated with 21 categories that differ in quality and quantity. Existing descriptions of the entire space rest on recordings made in the late 1990s or earlier, while recent work in general has focused on sub-sets of the space. The present paper reports two studies. The first presents static and dynamic acoustic analyses of the entire vowel space using a recently released database of *h-VOWEL-d* words (SwehVd). The second study compares the acoustic characteristics of SwehVd against formant data from 8 talkers of the same dialect recorded in 1999, to investigate the extent to which the phonetic realization of vowels has changed over the last generation. The results highlight the importance of static and dynamic spectral and temporal cues for Swedish vowel category distinction, and indicate possible ongoing shifts in the high front part of the space

Keywords: vowels, category separability, formant dynamics, vowel change

Word count: X

The acoustic characteristics of Swedish vowels

1 Introduction

The Swedish vowel inventory consists of 21 categories that differ in spectral (formant frequencies) and temporal cues (duration). It forms a typologically rather complex space, characterized by a systematic quantity distinction resulting in 9 long and short vowel pairs, 3 different levels of lip-rounding, and contextually conditioned allophones to /ɛ/ and /ø/ in position before /r/ or any retroflex segments. Given the crowdedness of the space and resulting category overlap for some vowels, previous work has reported on the need to look beyond static point estimates of the two primary determinants to vowel identity cross-linguistically, the first two formants (F1 and F2, e.g., Joos, 1948; Ladefoged & Broadbent, 1957; Nearey & Assmann, 1986; Peterson, 1961), as well as the hypothesized importance of additional cues besides F1 and F2, such as the third formant (F3) for rounded vs. unrounded categories (e.g., Fant, 1959; Fant, Henningsson, & Stålhammar, 1969; Fujimura, 1967; Kuronen, 2000), duration for certain long-short vowel pairs (Behne, Czigler, & Sullivan, 1997), and formant dynamics for some front contrasts (Kuronen, 2000; Pelzer & Boersma, 2019).

This paper investigates the acoustic characteristics of modern-day Swedish vowels in two studies that aim to contribute to our understanding of language-specific and language-general patterns of vowel acoustics. The first study presents a comprehensive description of the primary acoustic cues to vowel identity, using a recently released database of *h-VOWEL-d* (short: hVd) words recorded by 44 male and female talkers of Swedish, the SwehVd database (Persson & Jaeger, 2023). The second study assesses whether the Swedish vowel space has been submitted to changes over the last generation, by comparing the acoustics of SwehVd against that of a reference material from 1999 (Eriksson, 2004). The variety investigated is Central Swedish, the regional standard variety of Swedish

spoken in an area around and beyond Stockholm (eastern Svealand) (Bruce, 2009; Elert, 1994; Riad, 2014).[1] Existing descriptions of the entire space of 21 vowels rest on recordings made more than 25 years ago (reported in, e.g., Engstrand, 1999; Kuronen, 2000; T. Leinonen, 2010; Riad, 2014). Two of the most recent studies are T. Leinonen (2010) and Kuronen (2000) (Table 1). The former is based on recordings obtained around 1999 of all vowels, of which four short vowels were omitted from analysis. It covers 98 rural locations in Sweden and Swedish-speaking parts of Finland, including reference talkers of Standard Swedish. The latter covers the entire vowel space but is based on recordings from 1981 (K. Leinonen, Pitkänen, & Vihanta, 1981). More recent work over the last two decades has focused on parts of the phonological space, such as the long vowels for diphthongization studies (Pelzer & Boersma, 2019), two vowels for merger studies (e.g., [ɵ] - [œ] in Wenner, 2010), or a single vowel, e.g., the damped [ɨː] (Schötz, Frid, & Löfqvist, 2011), see Table 1. These studies all provide detailed mappings of different parts of the space, and contribute important insights into the current state of as well as ongoing processes. However, given their focus on subsets of the space, a comprehensive acoustic mapping of the modern-day Central Swedish vowel space *in its entirety* is lacking. Given that there is some evidence that productions of minimal pairs can lead to enhanced contrasts (e.g., Schertz, 2013; Seyfarth, Buz, & Jaeger, 2016), how representative such subsets are for the vowel space as a whole, remains an open question. In addition, most previous studies differ in the materials used, in terms of the size of the database (e.g., number of talkers and repetitions per vowel), the demographics of talkers (e.g., male/female talkers, region of origin), and phonological contexts used for recording. For instance, the majority of previous work has either not held the phonetic context constant across vowels, or has investigated isolated vowel production out of context or in different CVC contexts (Table 1). This diversity restricts comparison across studies on Swedish, as well as cross-linguistically.

[1] For the reader unfamiliar with Central Swedish, Section 1.1 provides an overview of the acoustics of Central Swedish vowel space.

Table 1
A selection of previous studies on Central Swedish vowels

Article	Speech materials	Participants	Approach	Focus
Eklund & Traunmüller, 1997	3 repetitions of all 9 long vowels in isolation	12 talkers (6 female) from the Greater Stockholm area, 20-58 years of age	Formant trajectories at 10 measurement points; F1-F2 means and SD; linear regression	Comparing the acoustics and perception of whispered and phonated vowels
Elert, 1964	1 repetition of all 21 vowels in sentence lists and word list (word list recorded by 2 talkers only; different V:C and VC: contexts)	11 talkers (5 female) from Stockholm, born in the 1930's	Duration means, SD and SE; t-statistics	Phonological quantity
Eriksson, 2004	3-4 repetitions of all 21 vowels, different phonological contexts for each vowel	12 talkers (6 female; 6 young) each from 107 locations across Sweden and Finland; 12 reference talkers (6 female; 6 young) of standard Swedish from the Greater Stockholm area	Project description	SweDia dialect database development
Fant, Henningsson & Stålhammar, 1969	1 repetition of all 9 long vowels in isolation	24 male talkers, students at KTH in Stockholm	F1 to F4 and duration means	Formant frequencies of the long vowels
Kuronen, 2000	5-7 repetitions of all 21 vowels, different phonological contexts for each vowel and each repetition	4 male talkers from Nyköping, 17-18 years of age (4 female reference talkers), recorded by Leinonen, Pitkänen & Vihanta, 1982	F0, F1 to F4 and duration means; formant trajectories at 4 measurement points with 30 ms intervals (long vowels)	Spectral and temporal acoustics of all Central Swedish vowels
Leinonen, 2010	a subset of 19 vowels from the SweDia database (see Eriksson, 2004 above)	see Eriksson, 2004	Principal component analysis of Bark-filtered vowel spectra; multidimensional scaling for dialectometric analysis	Dialectal variation in Swedish vowel pronunciation
Lindblom, 1963	5 repetitions of all 8 short vowels in bVb, dVd, gVg contexts, with four different stress patterns	1 male talker from Stockholm, 19 years of age	F1 to F3 means, duration; formants as a function of duration and consonsantal context	The effect of vowel duration on formant frequencies
McAllister, Lubker & Carlson, 1974	10 repetitions of all 6 long rounded vowels in itV context	6 talkers of standard Swedish	EMG; F1 to F3 means	The articulation and acoustics of rounded vowels
Pelzer & Boersma, 2019	8 repetitions of all 9 long vowels, different phonological contexts for each vowel and each repetition	8 talkers (4 female) from Stockholm	F1-F2 median values at 20, 50, 80% into the vowel; linear mixed effects model	Diphthongization in the long vowels
Schötz, Frid & Löfqvist, 2011	6 repetitions of [ɨː] in two different contexts (bibel, papipa)	1 male talker from Stockholm	Articulography; F1 to F4 means; Bark-circles	The articulation and acoustics of the damped [ɨː]
Wenner, 2010	3 repetitions of [œ] and [ɵ] in different phonological contexts	78 talkers (40 female) from 4 locations in Uppland, 12-85 years of age	F1 to F3 means; linear regression; correlation analysis	The merger of [œ] and [ɵ]

The materials and methodological approach adopted in the current paper is motivated by the goal to complement previous work for a comprehensive picture of modern-day Central Swedish vowels. The first study provides an up-to-date acoustic description of the entire vowel space of 21 categories, using the SwehVd database (Persson & Jaeger, 2023). The use of an hVd database minimizes coarticulatory effects from the surrounding phonetic context, and increases cross-linguistic comparability (e.g., Hillenbrand, Getty, Clark, & Wheeler, 1995; Peterson & Barney, 1952). The main spectral and temporal cues to vowel identity are reported in static analyses (following e.g., Engstrand, 1999; Fant et al., 1969), as well as dynamic analyses, given the well documented importance of formant dynamics on vowel production and perception (e.g., Assmann & Katz, 2005; Eklund & Traunmüller, 1997; Hillenbrand & Nearey, 1999; Kuronen, 2000; Nearey & Assmann, 1986). The static analysis assesses what cues contribute to vowel distinctions and evaluates some of the claims introduced in previous work, such as the hypothesized importance of F3 for rounded vs. unrounded high front contrasts (Fant, 1959; Fant et al., 1969; Fujimura, 1967; Kuronen, 2000), and to what extent spectral and temporal cues contribute to long-short vowel pair distinctions (e.g., Behne et al., 1997; Kuronen, 2000). The dynamic analysis explores what part of the space seems more prone to diphthongization, and investigates how formant dynamics contribute to vowel distinctions. In contrast with previous work investigating the dynamics of Central Swedish vowels, the present study includes both long and short vowels, thus submitting the entire vowel space to the same analyses.

Anticipating some of the main results from Study 1, the acoustic analyses suggested differences in the phonetic realization of some vowels compared to other qualitative descriptions of Central Swedish (Eklund & Traunmüller, 1997; Engstrand, 1999; Fant et al., 1969; Kuronen, 2000; Riad, 2014). These included the fronting of [eː], the lowering of [ɛː], and the centralization of [iː] and [yː] (c.f., T. Leinonen, 2010; Pelzer & Boersma, 2019; Schötz et al., 2011). This led me to conduct Study 2, which aimed to assess the scope and spread of these changes over the last generation, by comparing the acoustic characteristics

of SwehVd against vowel data from a database of 8 (female = 4) talkers of Central Swedish recorded in 1999 as part of the SweDia dialect project (Eriksson, 2004).

The paper is organized as follows. A background to the acoustics of Central Swedish vowels is provided by a review of previous work. This is followed by the two studies' methods and results, and finally, a discussion of the results and its consequences for the Central Swedish vowel system. All analyses and visualization code for this study can be found in an online repository (https://osf.io/7uvj4/). This article is written in R Markdown, which allows readers to easily replicate the analyses using freely available software (R Core Team, 2023; RStudio Team, 2020).

1.1 The acoustics of Central Swedish vowels

This section provides a description of the overall inventory of Central Swedish monophthongs, and discusses the role of cues beyond F1 and F2. It furthermore presents a review of previous studies on diphthongization and formant dynamics.

Central Swedish is most often described as having nine vowel phonemes: /i/, /y/, /ʉ/, /e/, /ɛ/, /ø/, /ɑ/, /o/, /u/. The long allophones are [iː], [yː], [ʉː], [eː], [ɛː], [øː], [ɑː], [oː], [uː], and the short allophones are [ɪ], [ʏ], [ɵ], [ɛ], [ø], [a], [ɔ], [ʊ]. The short allophones of /e/ and /ɛ/ has been reported to neutralize as [ɛ] in Central Swedish, resulting in 17 vowels, rather than 18 (Riad, 2014). There is also evidence of neutralization of the short /ø/ and /ʉ/ as [ɵ] among some talkers, primarily in position before a retroflex (Ståhle, 1965; Wenner, 2010). In addition to these 17 vowels, there are 4 additional long and short allophones—[æː], [æ], [œː], and [œ], as /ɛ/ and /ø/ lower in position before /r/ or any retroflex segment (e.g., Kuronen, 2000; Riad, 2014). Traditionally, Central Swedish has been described using four height levels and three backness levels (Riad, 2014).

It has furthermore been suggested that Central Swedish is defined by three levels of lip-rounding, where the rounded vowels are most often referred to as either inrounded, with

an extreme narrowing of the lips—[ʉː] and [uː], or outrounded, with a lesser degree of lip-narrowing and more protruded lips—[yː], [øː], [œː], [oː], and the remaining vowels defined as unrounded (e.g., Fant, 1971; McAllister, Lubker, & Carlson, 1974). Previous work has claimed that lip-rounding is particularly important for some vowel distinctions. For instance, [iː] and [yː] have been described as overlapping in F1-F2 space, but as more separable when F3 is considered (Fant, 1959; Fant et al., 1969; Fujimura, 1967; Kuronen, 2000).

The vowels in each pair have been reported to differ systematically in duration, with short-long vowel to vowel ratios on average .65-.67 for Central Swedish (Elert, 1964; Kuronen, 2000; Strangert, 2001). Spectral differences have traditionally been interpreted as a consequence of the durational distinction, hence assuming a trading relationship between spectral and temporal cues (for a review, see Schaeffler, 2005). It has been hypothesized that most of the durational variation is carried by F2 (e.g., Kuronen, 2000; Lindblom, 1963). Previous work has found the largest spectral differences for the [ʉː] - [ʊ], and [ɑː] - [a] vowel pairs, and the smallest differences for [ɛː] - [ɛ], and [øː] - [ø] (e.g., Kuronen, 2000). For pairs with small spectral differences, duration is presumably more important for vowel distinction. Perceptual studies on synthesized speech from talkers of Stockholm Swedish have confirmed that duration is the primary cue for [iː] - [ɪ], and [oː] - [ɔ] (Behne et al., 1997; for results on Southern Swedish and additional vowel pairs, [ɛː] - [ɛ], [øː] - [ø], see Hadding-Koch & Abramson, 1964). The extent to which *all* long-short vowel pairs rely on spectral cues is less known, as studies have focused on subsets of pairs.

According to previous work, several of the long vowels in Central Swedish tend to diphthongize in their phonetic realization. Diphthongization is considered prosodically conditioned and is the strongest in stressed vowels (Bleckert, 1987; T. Leinonen, 2010).[2] Previous studies have characterized the diphthongal glide in the later part of the long

[2] In general, true phonological diphthongs are not considered part of the phonological inventory of Central Swedish (Eliasson, 2022).

vowels as either a centralization of the vowel segment towards [ə] or a more open quality, or as a consonantal offglide (e.g., Elert, 1981, 2000; Fant, 1971; Fant et al., 1969; Kuronen, 2000; McAllister et al., 1974; Pelzer & Boersma, 2019; Riad, 2014). Results are inconclusive as to how widespread diphthongization is across the vowel space, and what direction it takes. Most work has however found substantial diphthongization towards a more open quality for the mid and mid-high vowels [eː], [øː] and [oː] (Eklund & Traunmüller, 1997; Elert, 2000; Fant et al., 1969; Pelzer & Boersma, 2019). In addition, diphthongization has been hypothesized to cue vowel distinctions for certain high vowels ([iː], [yː], [ʉː], and [uː]) (e.g., Fant, 1971; Kuronen, 2000). For instance, Kuronen (2000) reported that [iː] - [yː] - [eː], and [uː] - [oː], differed in formant patterns only at later time-points of the vowel for some talkers and that the contrast between [ɛː] and [æː] was maintained solely by trajectory movements. Of importance for the present study, less is known about the formant dynamics in the short vowels, given the almost exclusive focus on the long vowels in diphthongization studies.

Some talkers of Central Swedish have been reported to realize [iː], [yː], [ʉː] and [uː] with a consonantal offglide, where the end-point of [iː] is described as a palatal approximant [j], the end-point of [yː] a voiced labio-palatal approximant [ɥ], the end-point of [ʉː] and [uː] a voiced bilabial fricative [β] (Elert, 1980; Hammarström & Norman, 1957; McAllister et al., 1974). Furthermore, both [iː] and [yː] can be damped and produced with a buzzing sound, phonetically realized as [ɨː]. The damped [ɨː] has been found in several dialects across Sweden, both in rural areas and in the cities of Gothenburg and Stockholm (Björsten & Engstrand, 1999; Elert, 1980; Engstrand, Björsten, Lindblom, Bruce, & Eriksson, 2000; Gross & Forsberg, 2020; Riad, 2014). In work on Swedish dialectology, it is often referred to as the Viby-*i*, and in the Stockholm area, as the Lidingö-*i*. Acoustically, it manifests primarily as a lowering of F2, thus occupying a more centralized position in the F1-F2 space. Schötz et al. (2011) describe it as a central palatal vowel, as the articulatory correlates involve a retracted and lower tongue position, the tip of the tongue being higher

than blade and dorsum.

2 Study 1: Static and dynamic analyses of the spectral and temporal properties of Central Swedish

Study 1 describes the spectral acoustics of Central Swedish in static and dynamic analyses of F0, F1, F2, F3, and duration. It aims to provide a detailed description of the acoustics of all Central Swedish vowels, and to evaluate the relative importance of certain cues for specific vowel contrasts, as hypothesized in previous work. These include the importance of lip-rounding (F3) for high vowel distinctions (Fant, 1959; Fant et al., 1969; Fujimura, 1967; Kuronen, 2000), to what extent all long-short vowel pairs differ in quality (formants) and quantity (duration) (e.g., Behne et al., 1997; Kuronen, 2000), and what vowels seem to undergo diphthongization (Kuronen, 2000; Pelzer & Boersma, 2019). The dynamic analysis furthermore explores which cues carry information about neighboring vowel distinctions once dynamic information is considered.

The methodology employed in Study 1 is presented next, beginning with a description of the materials used.

2.1 Methods

2.1.1 Materials

The materials used in Study 1 is a corpus of Swedish hVd word recordings, collected by Anna Persson and Maryann Tan (Stockholm University) in 2020-2024, the SwehVd. An initial version of the corpus with 24 female talkers is described in Persson and Jaeger (2023). For this paper, an updated release is presented, including 20 additional male talkers (targeted number of male talkers = 24). All recordings, annotations, and acoustic measurements are available at https://osf.io/ruxnb/. SwehVd covers the entire

monophthong inventory of Central Swedish, including all nine long vowels, eight short vowels, and the four allophones to /ɛ/ and /ø/.[3] SwehVd focuses on a single regional variety, providing high resolution within and across talkers for this variety with N = 10 recordings of each hVd word from each of the N = 44 talkers (N = 24 female), for a total N of tokens = 9103. All talkers in the database were L1 talkers of Swedish, born and raised in the Greater Stockholm area or surroundings, of 18-44 years of age (mean age = 30; SD = 6.82). For more details on the recruitment, recording, pre-processing, segmentation and annotation procedure, see Persson and Jaeger (2023).

For the vast majority of talkers in the SwehVd, *hädd* productions elicited the same vowel as *hedd* (see Supplementary Information—SI, Figure S1), which confirms the commonly held assumption that the short allophone of /e/ neutralizes with the short allophone of /ɛ/ in Central Swedish. In order to have a balanced number of tokens for each vowel, all *hädd* words were excluded from the subsetted SwehVd materials used in this study (following Persson & Jaeger, 2023). Recordings on which the talker did not produce the targeted vowel were also excluded.[4] Furthermore, outliers were identified and removed by estimating the relative probability of each token's F1-F2 values given the joint distribution of F1-F2 for that vowel and talker. Tokens outside of the 2.50th to 97.50th quantile of the bivariate Gaussian distribution were filtered out. To facilitate empirical analyses and statistical models, all talkers (N = 7) with fewer than 4 remaining recordings for at least one of the vowels were removed. This left data from 37 L1 talkers (N=20 female talkers), with on average 322 (SD = 20) tokens per vowel (range = 277 to 345), for a total of 6759 observations.

---

[3] The words used to elicit the 21 vowels were: *hid*-[iː], *hyd*-[yː], *hud*-[ʉː], *hed*-[eː], *häd*-[ɛː], *höd*-[øː], *had*-[ɑː], *håd*-[oː], *hod*-[uː], *hidd*-[ɪ], *hydd*-[ʏ], *hudd*-[ɵ], *hedd*-[ɛ], *hädd*-[ɛ], *hödd*-[ø], *hadd*-[a], *hådd*-[ɔ], *hodd*-[ʊ], *härd*-[æː], *härr*-[æ], *hörd*-[œː], *hörr*-[œ].

[4] The SwehVd database contains information on both targeted vowel and what vowel was actually produced.

2.1.2 Acoustic analyses

2.1.2.1 Measuring acoustic cues to vowel identity The Swedish version of the Montreal Forced Aligner developed by Young and McGarrah (2021) was used to obtain estimates of word and segment boundaries. The boundaries were then manually corrected by the author (an L1-talker of Swedish). The formant analysis was carried out in Praat (Boersma & Weenink, 2022), using the Burg algorithm to extract estimates of the first three formants (F1-F3) at five time-points of the vowel (20, 35, 50, 65, 80% into the vowel), while vowel duration and F0 were extracted across the entire vowel segment. The Burg algorithm was parameterized with a time step of 0.01 seconds, a window length of 0.025 seconds, and pre-emphasis was applied from 50 Hz. The maximum number of formants was set to 5, with a formant ceiling of 5500 Hz for the female talkers, and 5000 Hz for the male talkers.

2.1.2.2 Vowel normalization The raw formant values were transformed into a vowel normalized space using Nearey's uniform scaling account (Nearey, 1978). Formant measurements in Hertz are reported in the SI, Section S1.2.4. Vowel normalization is used in studies on vowel production and perception to account for acoustically irrelevant inter-talker variation, as caused by differences in anatomical structure, e.g., vocal tract size (for reviews see e.g., Barreda & Nearey, 2018; Johnson & Sjerps, 2021; Stilp, 2020). In vowel production studies such as the present, normalization is primarily used as a methodological tool. Transforming the formant data into a normalized space reduces differences in F1 and F2 due to physiology, which can reduce between-talker variability and increase category separability, as visualized in Figure 1 (compare left and right panel).

Previous work on Swedish has primarily analyzed vowel data in raw Hertz (Björsten & Engstrand, 1999; Fant et al., 1969; Pelzer & Boersma, 2019), or transformed into Bark (Fant, 1983; Kuronen, 2000; Schötz et al., 2011; Wenner, 2010), Mel (Lindblom, 1963), or Lobanov (Gross & Forsberg, 2020). The choice of Nearey's uniform scaling in the present study was motivated by its previous use in socio-phonetic research to describe and compare

languages and varieties (e.g., Barreda, 2021; Labov, 2001; Labov, Ash, & Boberg, 2005), and by its plausibility as perceptual model of how we come to achieve robust cross-talker perception, as it has provided a good fit against both production (e.g., Persson & Jaeger, 2023; Syrdal, 1985) and perception data (e.g., Barreda, 2021; Persson, Barreda, & Jaeger, 2024).

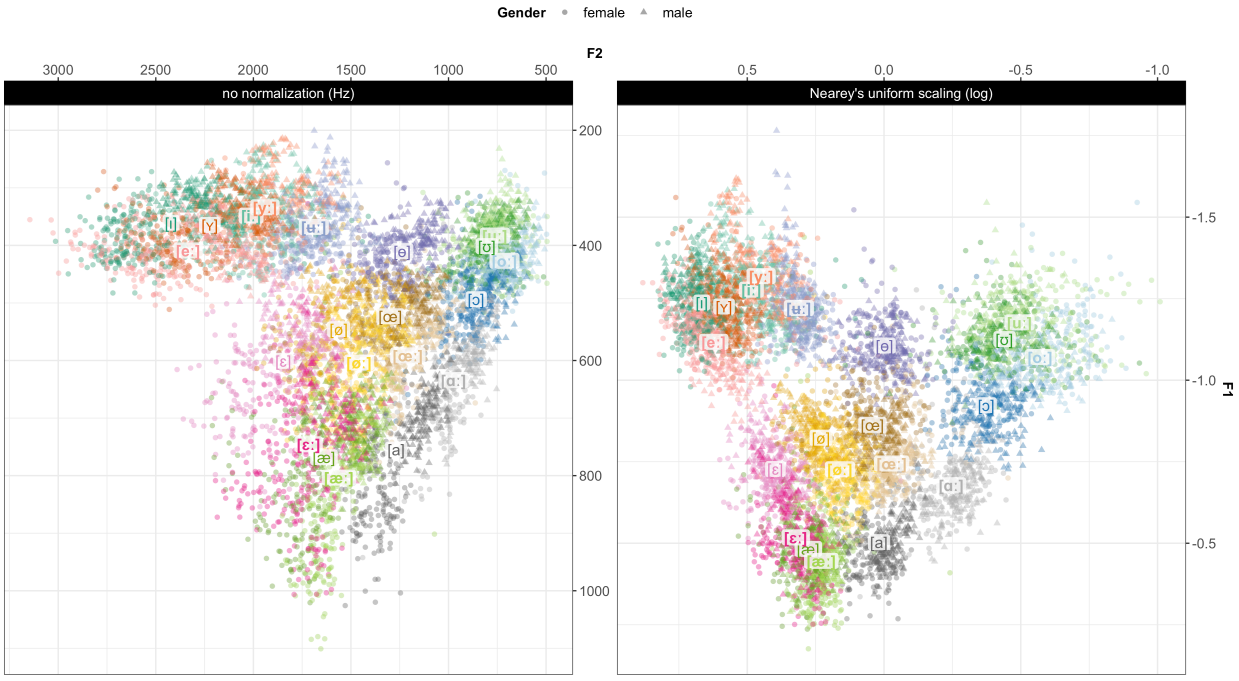

Figure 1. The SwehVd vowel data in unnormalized Hertz (*left*) and Nearey's uniform scaling space (*right*), along the first two formants, F1 and F2. Points show recordings of each of the 21 Central Swedish vowels by 44 (24 female) L1 talkers in the database, averaged across the three middle time-points (at 35, 50, 65% into the vowel). Vowel labels are placed at the vowel mean across talkers. Long vowels are boldfaced. Vowels that mismatched intended label are excluded (1.33% of all recordings).

2.1.2.3 Static acoustic analysis The static analysis of SwehVd presents formant measurements at the steady state of the vowel, by averaging across the three mid-points.[5]

[5] The choice of time-point for extracting formants, or whether to average across several time-points, affects the acoustic characterizations given how formants move across the vowel segment. The SI (S1.2.2) presents evaluations of the effect of different measurement points. For increased comparability across studies in this paper, the steady state of the vowel was selected for both studies.

It maps the entire vowel space of 21 categories and evaluates the relative contribution of F0, F1, F2, F3 and duration to vowel distinctions, using visualizations of cues and cue correlations. While fundamental frequency (F0) is not considered an important cue to vowel identity in itself, it is known to vary between languages, dialects and speech styles (e.g., Henton, 2005; Jacewicz & Fox, 2018; Johnson, 2005; Mennen, Schaeffler, & Docherty, 2012; Weirich, Simpson, Öjbro, & Ericsdotter Nordgren, 2019) and is therefore reported.

In order to evaluate the hypothesized importance of lip-rounding (F3) for neighboring unrounded and rounded categories, a category separability index was employed. Following work by Wedel, Nelson, and Sharp (2018) and X. Xie and Jaeger (2020), each vowel's separability from the neighboring vowel was calculated as the average distance of vowel tokens to the centroid of the neighboring vowel, operationalized as (1).

$$separability\ of /\text{yː}/ from /\text{iː}/ = \frac{\sum_{k=1}^{n} \sqrt{(F1_{token\ k\ of /\text{yː}/} - F1_{Center\ of /\text{iː}/})^2 + (F2_{token\ k\ of /\text{yː}/} - F2_{Center\ of /\text{iː}/})^2}}{n} \quad (1)$$

For instance, for the [yː] - [iː] contrast, first, each talker's [iː] center was calculated for F1-F2. Next, the distances between each [yː] token to the neighboring [iː] center from the same talker were calculated for F1-F2. Finally, the distances were averaged across all [yː] tokens from a talker, resulting in a separability measure for that vowel and talker. The higher the index, the greater the separation between categories. The same was subsequently done for F1-F2-F3. These two measures of separability for each contrast (F1-F2, F1-F2-F3) were then compared to assess whether including F3 would lead to increased category separability. The contrasts investigated were [yː] - [iː], [eː] - [yː], [ɪ] - [ʏ] for comparing unrounded vs. outrounded vowels, and [yː] - [ʉː], [oː] - [uː], [ɔ] - [ʊ] for outrounded vs. inrounded vowels.

To quantify the effect of including F3 on category separability, separate linear mixed-effects model (LMM) were fit for each contrast, predicting separability from cue

combination (F1-F2-F3 vs. F1-F2) while including by talker random intercepts.[6] The model was formulated as follows: $separability \sim cuecombination + (1|Talker)$. Cue combination was sum-coded (F1-F2 = 1, F1-F2-F3 = -1).

The same process was applied to investigate to what extent long-short vowel pairs differ in spectral cues, by assessing what combination of spectral cues could provide the largest separability between the two vowels in each pair. For this evaluation of quantity contrasts, the category separability index was calculated for each pair and four different cue combinations: F1-F2, F1-F3, F2-F3 or F1-F2-F3. The models were the same as the previous sets, however, cue combination was treatment-coded with F1-F2 as reference category, thus comparing each cue combination against the F1-F2 combination.

The results of the static analysis are presented in Section 2.2.1.

2.1.2.4 Dynamic acoustic analysis Formant measurements at all five time-points were used in the dynamic analysis to assess the importance of formant dynamics for vowel distinctions. The dynamic analysis is divided into two main sections. In the first section, formant trajectory plots were used to assess the scope and direction of formant movements, to what extent vowels seemed to diphthongize, and to evaluate the hypothesized importance of formant trajectories for the [iː]-[yː]-[eː], [oː]-[uː] and [ɛː]-[æː] contrasts reported in previous work (e.g., Kuronen, 2000; Pelzer & Boersma, 2019). Lastly, trajectories of short vowels were also visualized as they have not been typically explored in the past.

In the second part of the dynamic analysis, the hypothesized contribution of formant dynamics to category information was modeled using generalized additive mixed-effects models (GAMMs) (Baayen, Vasishth, Kliegl, & Bates, 2017). GAMMs were employed to assess what cues carry information about vowel quality once formant dynamics were inspected. GAMMs are increasingly used in phonetic research, due to their suitability in modeling the non-monotonic complex phonetic patterns found in formants without

[6] By-talker random intercepts was the maximum random effect structure that converged.

assuming linearity or having to rely on the simplifying assumption that vowels can be reduced to a single F1-F2 point estimate (e.g., Chuang, Fon, Papakyritsis, & Baayen, 2021; Sóskuthy, 2021; Wieling, 2018). GAMMs have been used in studies on vowels in different English varieties, e.g., on /u/-fronting in Derby English (Sóskuthy, Foulkes, Hughes, & Haddican, 2018) and on the front vowel system of Southern American English (Renwick & Stanley, 2020) but to the best of my knowledge, they have not been implemented in studies of Swedish vowels. The use of GAMMs thus complements previous work on Central Swedish that has primarily used visual inspection, formant measurements and linear models (Table 1).

Two main groups of GAMMs were fit in the dynamic analysis. In the first group, GAMMs were fit to 4 subsets of neighboring contrasts, hypothesized to differ primarily in formant dynamics: [iː] - [yː] - [ʉː] - [eː], [oː] - [uː], [ɛː] - [æː] (Fant, 1971; Kuronen, 2000; Pelzer & Boersma, 2019). Given the directionality in formant trajectories found for [øː]-[œː] (Figure 7), this contrast was also included. To explore potential effects of dynamics in the corresponding short vowels, an additional 4 contrasts were modeled. These were not entirely identical to the long subsets, for reasons of evident separability in F1-F2 space: [ɪ] - [ʏ], [ɔ] - [ʊ], [ɛ] - [æ] and [ø]-[œ]. The general model formulation was as follows $formant \sim category + Gender + s(timepoint, by = category, k = 5) + s(Talker, bs = "re") + s(Talker, category, bs = "re")$. The GAMMs were treatment coded with [iː], [oː], [ɛː], [ɪ], [ɔ], [ɛ], and [ø] as reference categories in respective set.

The second group consisted of GAMMs fit to all 21 categories, aiming for an evaluation of differences between categories within vowel pairs. Vowel was backwards difference coded, with [iː] as reference vowel. The vowel variable was ordered alternating long and short vowels by the pair, so the short vowels were compared against their preceding long counterpart.

All GAMMs were fit separately for each of the formants, which necessarily meant committing to the simplifying assumption of cue independence. Previous work has shown

that acoustic cues tend to co-vary (for a review, see Schertz & Clare, 2020). For vowels, this is the case for F1 and F2, as shown by the shape and orientation of ellipses in Figure 2.

The results of the dynamic analysis are presented in Section 2.2.2.

2.2 Results

2.2.1 Static spectral and temporal cues to vowel identity

The static analysis begins with a mapping of the entire 21 category space along F1-F2. Next, the relative contribution of additional cues beyond F1 and F2 is assessed, as well as the extent to which all long-short vowel pairs are qualitatively and quantitatively different.

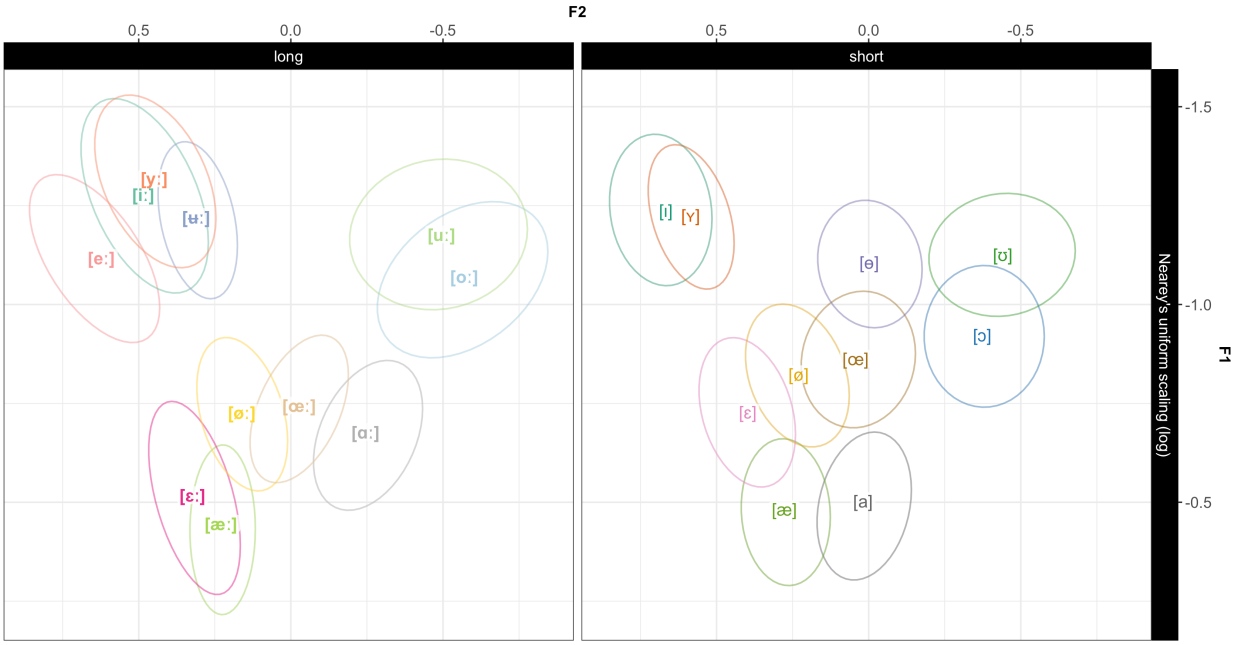

Figure 2. The SwehVd vowel data separated by quantity. Ellipses show bivariate Gaussian 95% confidence interval of vowel means. Vowel labels indicate vowel means across female and male talkers.

Figure 2, left panel, visualizes the long vowels along the two primary cues to vowel identity, F1-F2.[7] Four vowels cluster in the high front part of the space. The mid-high [eː]

[7] The SI presents the mean cue values for the male and female talkers, Tables S1 and S2. As expected, the

occupies a substantially higher position than in many previous descriptions, and is also the most fronted vowel (c.f., Fant et al., 1969; Kuronen, 2000; but see Engstrand et al., 2000; Pelzer & Boersma, 2019). The high [iː] and [yː] are rather mid-central, and exhibit substantial overlap with [ʉː]. The [uː] - [oː], and [ɛː] - [æː] categories are also partly overlapping.

The short vowels (right panel), present a slightly more compact space, however with increased category separability (c.f., Riad, 2014).[8] For some vowel pairs, overlap is clearly reduced for the short vowels, e.g., for [ɪ] - [ʏ], [ɛ] - [æ], and [ɔ] - [ʊ]. Of note, the high vowels [ɪ] and [ʏ] are more fronted than their long counterparts, which does not replicate previous studies on Central Swedish (e.g., Fant, 1971; Kuronen, 2000).

2.2.1.1 Cues and cue correlations For the pairwise combinations of the five spectral and temporal cues—F0, F1, F2, F3 and duration, see Figure 3 from Persson and Jaeger (2023) updated to include data from the 17 male talkers. Unsurprisingly, the densities along the diagonal suggest that F0 carries the least information about vowel identity, exhibiting less between-category separation than all other cues.

As is to be expected, vowels differing in quality are most separated in the F1-F2 panels. The F1-F3 and F3-F2 panels both display increased separation between the neighboring outrounded [yː] and inrounded [ʉː], and unrounded [ɪ] and outrounded [ʏ], compared to when plotted along F1-F2, which points to the importance of F3 for these vowels. Interestingly, the almost complete overlap between [iː] and [yː] in F1-F2 space overall remains when F3 is considered, even if some individual differences in the amount of overlap exist. Most talkers produce these two vowels very close in F1-F2 space, and only

male talkers have lower formant values and lower F0s than the female talkers (average F0 across long and short categories for female talkers = 203, for male talkers = 121).

[8] There is a possibility that the increased separability found for the short vowels is partly an artifact of how time-points for cue measurements were selected. Time-points based on percentage of vowel duration will necessarily render measurement points that are closer in time for the shorter vowels, potentially providing a better estimate of the formant value that is most distinctive, i.e., the steady state in the center of the vowel.

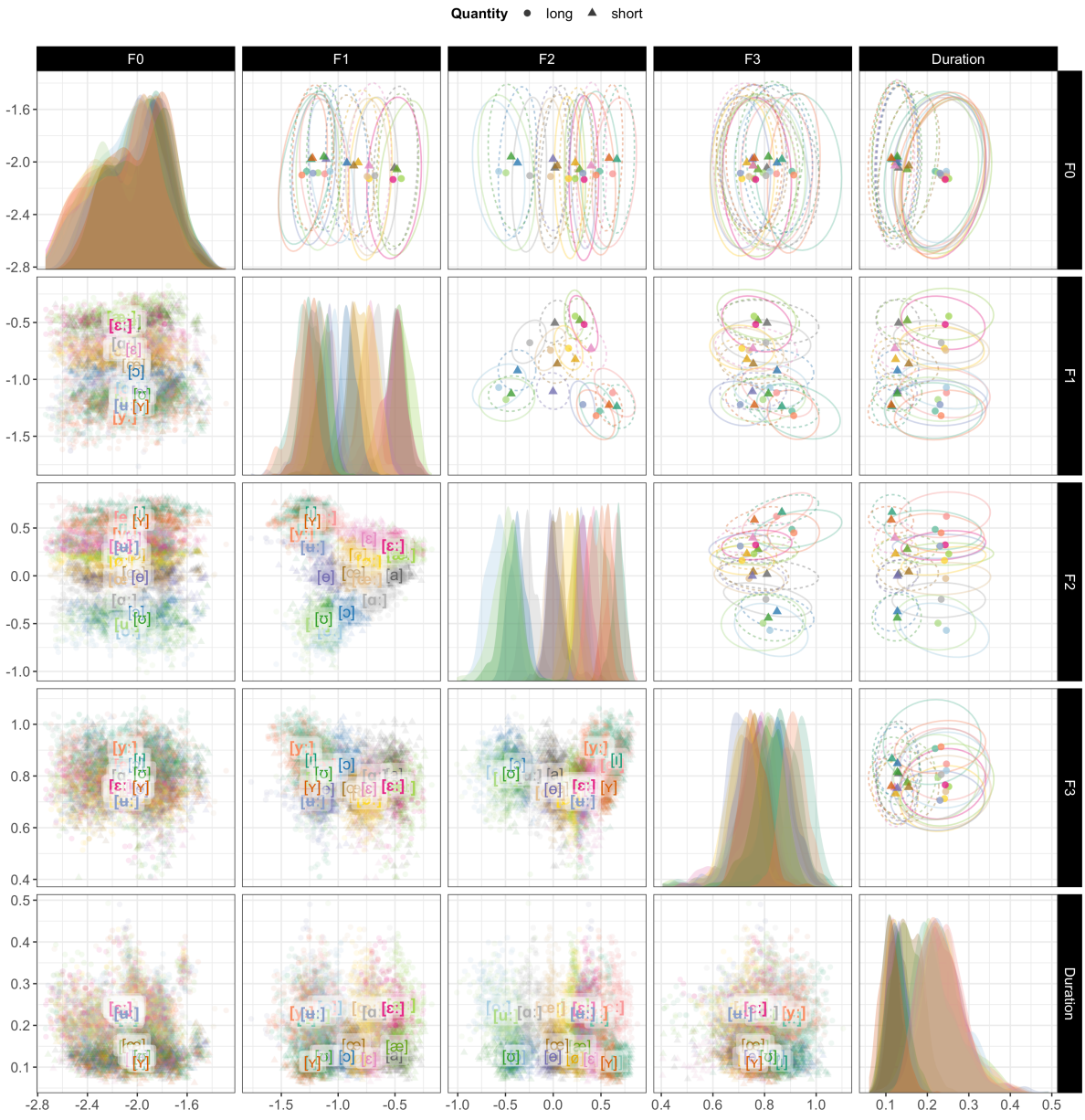

Figure 3. The SwehVd vowel data shown for all pairwise combinations of five cues: F0, F1, F2, F3 and duration. Panels on the diagonal show marginal cue densities of all five cues. The off-diagonal panels show vowel means across talkers, represented by points and with bivariate Gaussian 95% probability mass ellipses in the upper panels, and represented by vowel labels and with points for each recording in the lower panels. Note that, unlike in Figure 1, axis directions are not reversed.

slightly separated in F2-F3 space, while others display a continued overlap when considering F3 (for reference, one talker of each type are displayed in SI Figure S4). This would seem to suggest that F3 might carry less importance as distinctive feature for [iː] - [yː] than previously established (c.f., Fant, 1959; Fant et al., 1969).

In order to quantitatively asses whether the distinction between closely neighboring unrounded and rounded categories increased when F3 was considered, the category separability of these vowels was calculated based on F1 and F2, and subsequently compared against the separability calculated when including F3. If separability were to increase when F3 was added, it would suggest that F3 does contribute to category distinctions.[9]

Two general observations can be made from Figure 4. Category separability was overall lower for some contrasts when only F1 and F2 were considered, e.g., the [iː] - [yː], and [ɪ] - [ʏ] contrasts, presumably indicating their overlap in F1-F2 space. Second, including F3 increased overall category separability, but only marginally for most contrasts. The contrast that seem to benefit most from the inclusion of F3 is the [yː] - [ʉː] contrast.

The LMMs fit to the data (presented in Section 2.1.2.3) indicated that including F3 improved category separability for all contrasts (all $ps < .001$). This suggests that the subtle differences observed by visual inspection were nevertheless significant (Summary tables in SI S1.2.6).

2.2.1.2 Quantity vs. quality in long and short vowel pairs To gain more insight into the extent to which there are spectral and temporal differences between long and short vowels, the acoustics of categories within vowel pairs were evaluated. This allows for an assessment of whether quantity and quality distinctions seem to be separate from each other.

[9] To capture possible differences in separability between vowels in a contrast, the separability index was first calculated in both directions, that is, both assessing the separability of [iː] from [yː], and [yː] from [iː], following X. Xie and Jaeger (2020). Given that no significant differences in directionality were found, the separability index reports the average separability across the two categories in each contrast.

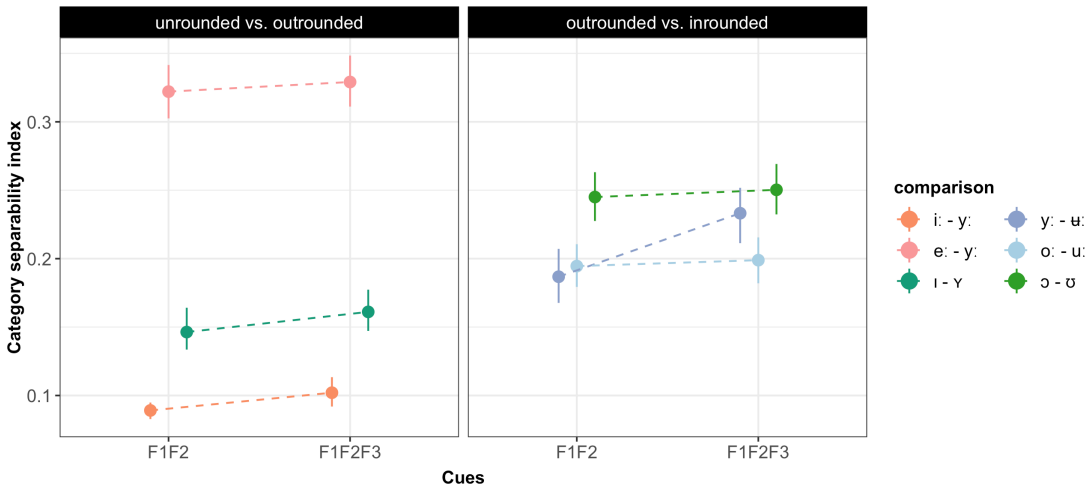

Figure 4. The effect of including F3 in measures of category separability for the distinction between neighboring unrounded vs. outrounded vowels (**left panel**), and outrounded vs. inrounded vowels (**right panel**). Pointranges indicate mean and 95% bootstrapped CIs of the category separability summarized across talkers for each cue combination.

As expected, long-short vowel pairs differ systematically in duration (Figure 5). For each vowel pair, the duration densities in Figure 5 are overlapping but with two clearly separable peaks (mean duration for the long vowels = 0.19 ms, SD = 0.10; mean duration for the short vowels = 0.08 ms, SD = 0.09). Overall, the short vowels display less variability in duration than the long vowels, a common pattern for measures with a lower bound.

All long-short vowel pairs furthermore display spectral differences in F1-F2. In fact, as indicated in Figure 1, formant differences are apparent for *all* vowel pairs, even for vowel distinctions for which duration has been found to be the primary cue—[ɛː] - [ɛ], [øː] - [ø], [iː] - [ɪ], and [oː] - [ɔ] (e.g., Behne et al., 1997; Hadding-Koch & Abramson, 1964; Kuronen, 2000). The vowel pairs that display larger spectral differences along F1-F2 seem to be [ʉː] - [ɵ] and [ɑː] - [a] (in line with e.g., Fant, 1983; Kuronen, 2000), but also [ɛː] - [ɛ], which contrasts with previous studies. The large spectral differences in [ɛː] - [ɛ] are presumably due to [ɛː] being produced very low in the SwehVd materials, which increases the distance

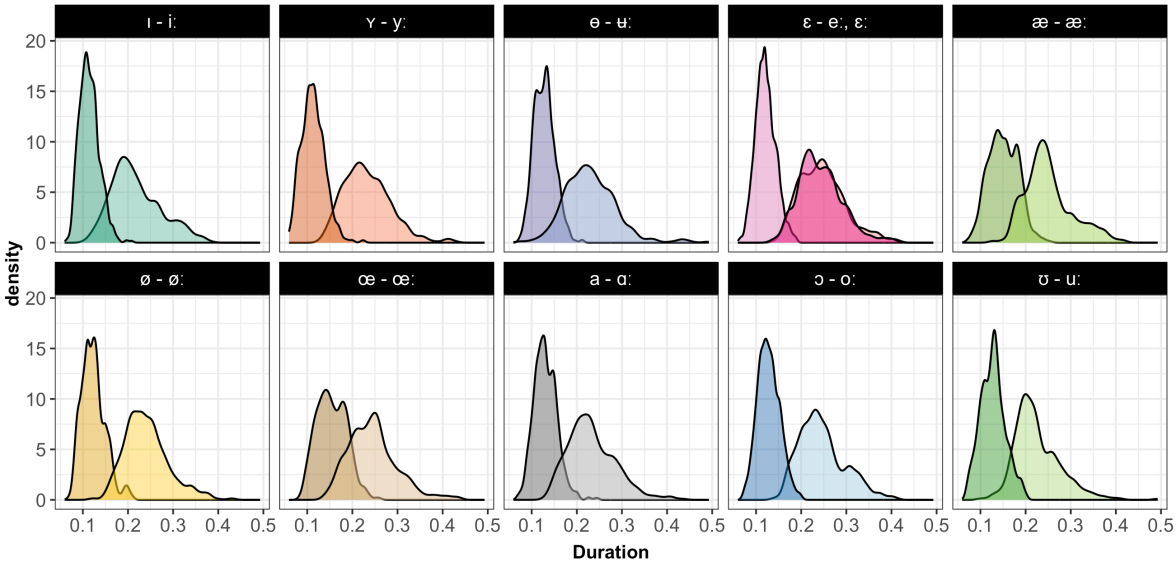

Figure 5. Illustrating the systematic differences in duration between the long and short vowel pairs in SwehVd.

to [ɛ] and in addition leads to a gap along the top-left to bottom-left diagonal between [eː] and [ɛ]. Overall, F2 appears to carry more of the spectral variation between the long and short vowel phonemes, as categories display increased separability in the pairwise combination of F2 and duration (Figure 3, rightmost column, third row).

In order to evaluate what cue combination would provide the largest separability between vowels in long-short contrasts, the category separability index was calculated for each pair and four different cue combinations: F1-F2, F1-F3, F2-F3 or F1-F2-F3.

Figure 6 indicates that category separability is generally higher for some pairs, e.g., [eː] - [ɛ], [ʉː] - [ɵ], and [ɑː] - [a], and that the F1-F2-F3 combination overall maximizes separability between the pairs, although the difference to the F1-F2 space appears marginal for most pairs. This would seem to suggest that the first two formants are the most important cues to long-short vowel pair distinctions, while the inclusion of an additional cue unsurprisingly does not punish the separability. For the F1-F3 and F2-F3 cue spaces, there seems to be two main groups for which either cue combination generates the lowest

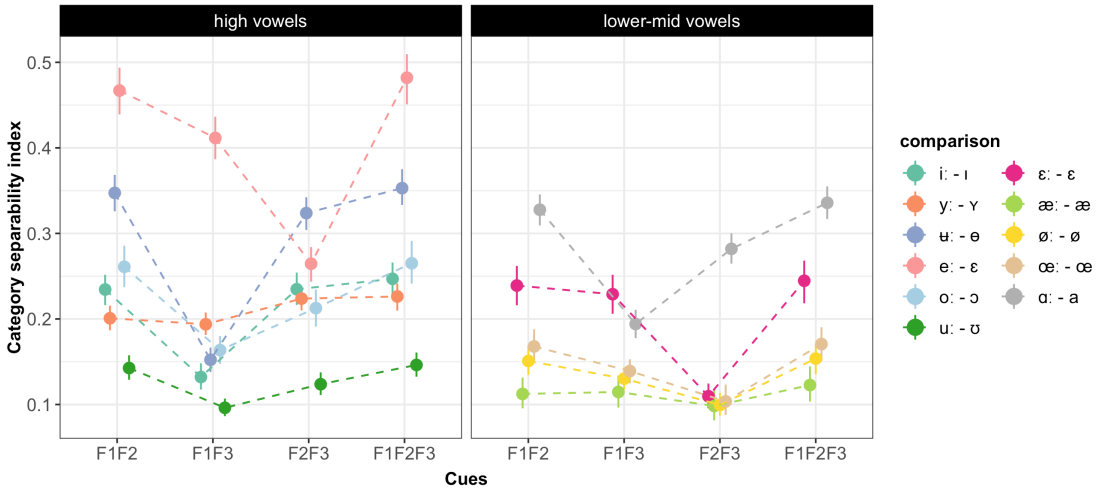

Figure 6. Category separability for long-short vowel pair distinctions depending on the cue combination considered. For visualization purposes, the pairs are split into high vowels (**left panel**), and lower-mid vowels (**right panel**). Pointranges indicate mean and 95% bootstrapped CIs of the category separability summarized across talkers for each cue dimension.

separability index. For the [ʉː] - [ɵ], [ɑː] - [a], [oː] - [ɔ], [iː] - [ɪ], and [uː] - [ʊ] vowel pairs, the F1-F3 cue combination generates the lowest separability, which would seem to suggest the informativity of F2 for distinguishing between these pairs. For the second group, mainly consisting of allophones to /ɛ/ and /ø/: [eː] - [ɛ], [ɛː] - [ɛ], [æː] - [æ], [øː] - [ø] and [œː] - [œ], the F2-F3 cue combination produces the lowest separability, presumably highlighting the importance of F1. Interestingly, the [yː] - [ʏ] pair generates almost identical separability indices in F1-F2 and F1-F3 spaces, and increased separability for both F2-F3 and F1-F2-F3, which seems to signal the importance of all three cues.

These findings were confirmed by the statistical analysis (Summary tables in SI S1.2.6): the LMMs indicated that the F1-F2-F3 cue combination generated the highest separability, however, for all but the [yː] - [ʏ] ($\hat{\beta} = .025, SE = .006, p < .0001$) and [eː] - [ɛ] ($\hat{\beta} = .015, SE = .007, p < .041$) vowel pairs, this change was significantly indistinguishable from the F1-F2 cue combination (all other $ps > .06$). For the other comparisons, F1-F2

vs. F1-F3, and F1-F2 vs. F2-F3, the LMMs indicated significant negative changes for both comparisons for all but three vowel pairs: for the [iː] - [ɪ], there was a significant effect of cue combination in the F1-F2 vs. F1-F3 comparison only ($\hat{\beta} = -1.02, SE = .007, p < .0001$), for the [yː] - [ʏ] and [æː] - [æ] vowel pairs, there was a significant effect in the F1-F2 vs. F2-F3 comparison only (for [yː] - [ʏ], $\hat{\beta} = .023, SE = .006, p < .0002$; for [æː] - [æ], $\hat{\beta} = -.014, SE = .005, p < .011$).

To sum up, the results of the static analysis suggests that F1 and F2 are the most important cues to vowel distinctions in Central Swedish. While visual inspection suggested that including F3 did not substantially increase category separability for neighboring rounded vs. unrounded contrasts, the statistical analysis found significant improvements in separability for all contrasts. This highlights subtle but significant differences, and the advantages of expanding empirical analyses to modeling approaches.

In addition, even though all long-short vowel pairs differed systematically in duration, they also displayed considerable spectral differences, suggesting that quantity distinctions—long vs. short vowels—are not separate from quality distinctions—high, low, front, back vowels. The comparison of how the category separability within each pair changed as a function of cue combination, furthermore highlighted the importance of F1 and F2, with F2 carrying much of the informativity for several pairs. The increase in separability found when including F3 was numerical for all pairs but the [yː] - [ʏ] and [eː] - [ɛ] pairs, indicating the importance of F3 for these two contrasts. Given that /y/ is a rounded category, the lower F3 values found in [ʏ] relative to [yː] would seem to indicate a relaxation of lip-rounding in [yː].

Given that the category separability index assigns equal weight to all cues included, there is no direct way of knowing which cue contributes more to the separability index. Furthermore, similar to other evaluations presented in this subsection, the separability index cannot account for the fact that formants are not static but rather fluctuate across the signal. A more holistic mapping of the acoustics should therefore aim to assess how

formant dynamics contribute to vowel distinctions. The next section investigates how formants move across the segment and how much information is gained by accounting for this dynamics.

2.2.2 Dynamic spectral analysis

This subsection begins with visualizations of the empirical data in formant trajectory plots. Next, the results of the GAMMs fit to the data are presented, first focusing on the movements in eight sets of neighboring vowel contrasts, then on spectral differences between the long and short vowel pairs.

2.2.2.1 Formant movements across the space Figure 7 displays the formant trajectories across all 5 time-points for the long and the short vowels. In almost all vowels, long or short, formants showed a dynamic pattern. Only [øː], [ɪ], and [ʏ] showed little movement over the measurement points. The scope and direction, however, vary. Across vowels, the scope of movements appear to be larger moving from vowel mid-point to 80% into the vowel, as indicated by the length of the line from vowel label to end of arrow. Most of the formant dynamics thus take place *after* vowel mid-point. The long high front vowels are important exceptions—the dynamics in [iː] and [yː] mostly occur at the beginning of the vowel segment (between 20 and 50% into the vowel), whereas [ʉː] displays movements of almost equal magnitude across the first four time-points. The largest movements overall seem to concern [eː] and [oː].

In terms of directionality, there is a general tendency to move towards the centre in most of the vowels, both long and short. According to previous studies (e.g., Bleckert, 1987; Elert, 2000), the high vowels [iː], [yː], [ʉː] and [uː] tend to be realized with an offglide, which would generate a falling F1 for all four vowels, a rising F2 for [iː] and [yː] and a falling F2 for [ʉː] and [uː]. These predictions were borne out for F1 in all cases, but for F2, only for [ʉː]. Both [iː] and [yː] display very little movement along F2, whereas [uː] moves towards a

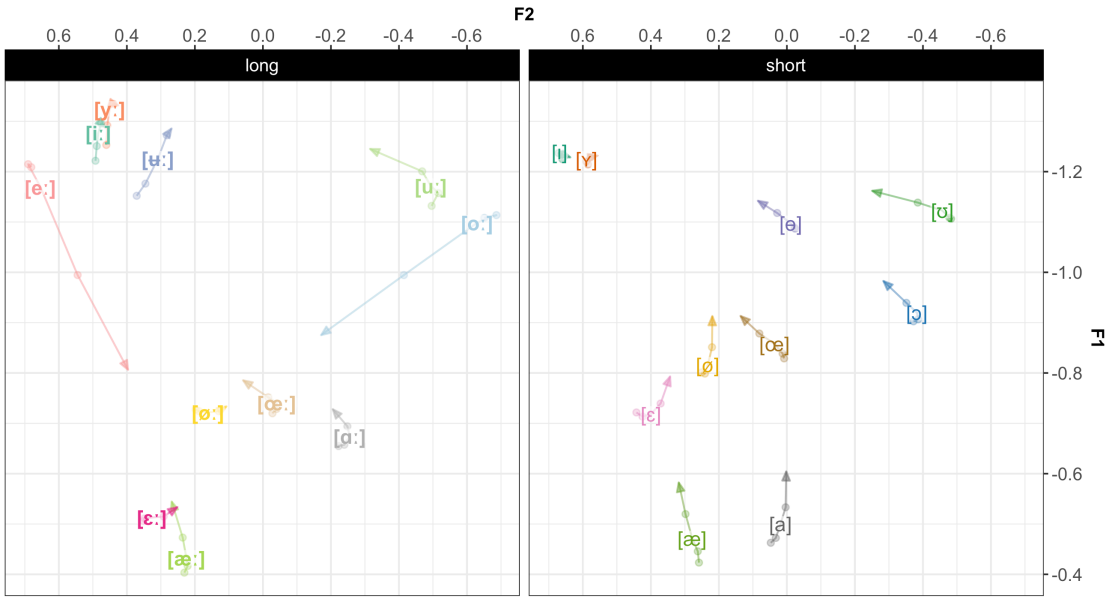

Figure 7. The trajectory of all vowels across the five time-points, along F1-F2. The arrow indicates the direction of the trajectory and ends at the final time-point, at 80% into the vowel. The vowel label is placed at the third time-point, at vowel mid-point (50%). The first (20%), second (35%) and forth (65%) time-points are represented by points.

more central quality, possibly indicating diphthongization ending in [ə] rather than a consonantal offglide.

Parts of the movements could be due to coarticulatory effects in anticipation of the upcoming coda ([d], [ɖ], [r]). If so, one would expect F2 to centralize in the later part of the segment, as tongue movements mark transitions into the alveolar (e.g., Hillenbrand, Clark, & Nearey, 2001; K. N. Stevens & House, 1963). The formant movements along F2 from the last point (65%) to arrow tip (80%) in e.g., [ɵ], [œː], [œ], [ɑː], [ɔ], [uː], and [ʊ], might at least partly be caused by such coarticulation. Given the scope and direction of movements, Figure 7 suggests diphthongization in primarily [eː], [ʉː] and [oː], replicating previous work (e.g., Eklund & Traunmüller, 1997; Elert, 2000; Pelzer & Boersma, 2019), while the other vowels appear to merely display formant movement, that partly could be caused by e.g., coarticulation. The previously reported diphthongization in [øː], however, does not seem to be particularly pronounced in these data.

Figure 7 further demonstrates that some neighbouring categories either converge at end points or diverge at end points. For instance, [uː] - [oː] are fairly closely located at earlier time-points, but differ substantially towards the end of the vowel segment, while [ɛː]-[æː], [øː]-[œː] and [ø]-[œ] start at different locations but end up in approximately the same (c.f., Kuronen, 2000). Finally, the formant trajectories suggest that the empty spots identified in the vowel space under a static analysis (Figure 1), may indeed be occupied when vowel dynamics are considered. This is especially true for [eː] that travels from the mid-high front to the mid center of the space as the signal unfolds, down to a position closer to its short counterpart, [ɛ]. Given the amount of overlap when static spectral cues are considered (Figure 2), formant dynamics are likely highly informative for several of these distinctions. Figure 8 parallels Figure 2 and illustrates the effect of considering formant movements for neighboring categories. As visualized in Figure 8, the overlap between [uː] - [oː] is substantially reduced at the later time-points, while [ɛː]-[æː], [øː]-[œː] and [ø]-[œ] are most distinguishable at earlier time-points.

2.2.2.2 Models of formant dynamics

2.2.2.2.1 The effect of modeling formant dynamics for neighboring contrasts

The first set of GAMMs were modeled separately for each cue (F1, F2, F3) and each of the sets of neighboring vowels hypothesized to (at least for some talkers) rely on formant dynamics (Fant, 1971; Kuronen, 2000; Pelzer & Boersma, 2019): the high front vowel contrasts [iː] - [yː] - [ʉː] - [eː] and its short counterpart [ɪ] - [ʏ], the high back vowel contrast [oː] - [uː] and its short counterpart [ɔ] - [ʊ], the lower-mid front contrast [ɛː] - [æː] and short [ɛ] - [æ], and the mid center [øː] - [œː] and [ø] - [œ]. Summary tables of models are included in the SI, Section S1.2.7.

The GAMMs fit to the high front vowel contrasts suggested an effect of vowel on all cues for all vowels (all $ps < .0025$), except for [yː] predicting F3 ($\hat{\beta} = .0007, SE = .013, p > .95$), and for [ʏ] predicting F1 ($\hat{\beta} = .01, SE = .01, p > .32$).

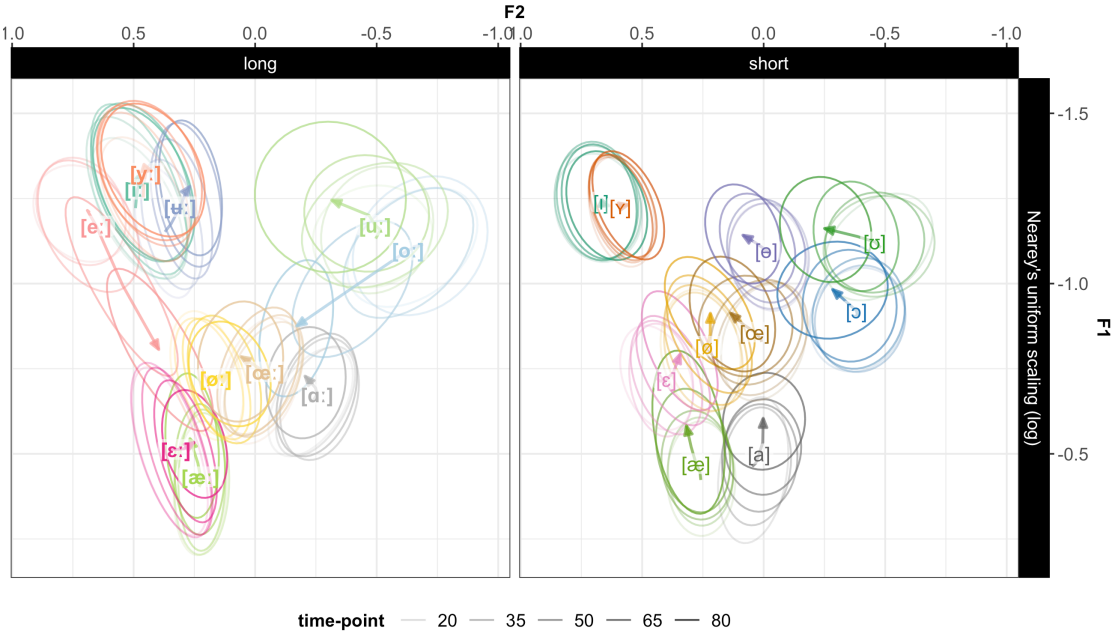

Figure 8. Vowel placement in F1-F2 space at each of the five time-points. Ellipses show bivariate Gaussian 95% confidence interval of vowel means at each of the five time-points. Transparency indicates time-point, more transparent ellipses for earlier times. The vowel label is placed at vowel mid-point (50%). The arrow indicates the direction of the formant trajectory and ends at the final time-point, at 80% into the vowel.

The long [yː] predicting F2 was also close to insignificant ($p > .046$). This indicates that F3-dynamics does not aid in distinguishing between [iː] and [yː], and neither does F1-dynamics for [ɪ] and [ʏ]. These results would thus overall seem to suggest that while F3 increases separability between [iː] and [yː] under static analysis, the *dynamics* in F3 does not seem to add anything to the contrast. For the [ɪ] - [ʏ] contrast, F3 is likely more informative, while the F1-dynamics does not contribute to distinguishing between the two. Figure 9 panels A and E demonstrate the importance of formant dynamics for the high front long vowels in F1 and F3, as many of the vowels are overlapping at some time-points, but never along the entire segment.

An effect of category on all three cues was found in all GAMMs fit to the high back

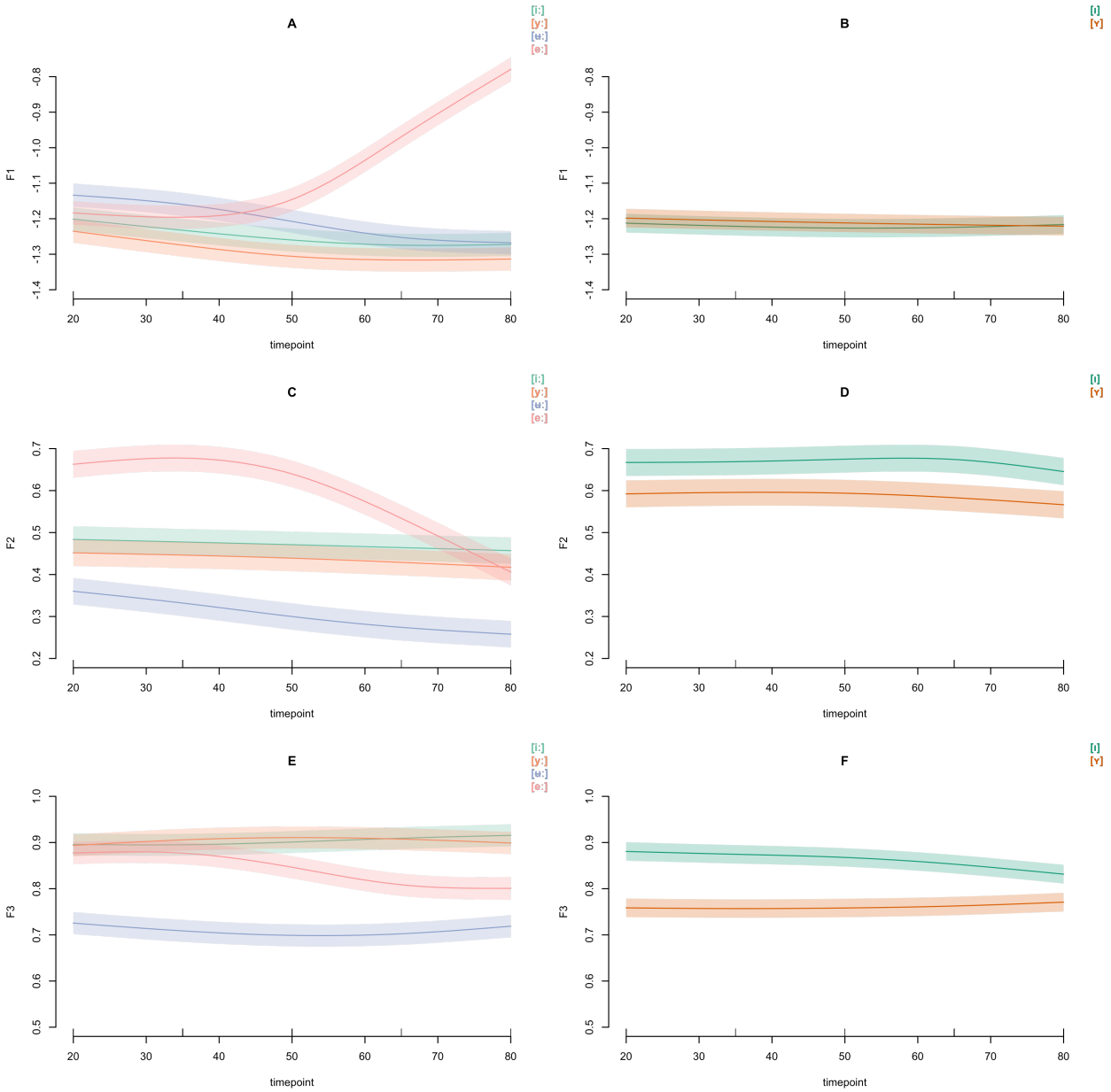

Figure 9. Fitted smooths of GAMM for predicting F1 (**upper row**), F2 (**mid row**), F3 (**bottom row**) and 95% confidence intervals for the long vowels [iː], [yː], [ʉː], and [eː] (**left**), and the short vowels [ɪ] and [ʏ] (**right**).

long and short contrasts (all $ps < .0006$). These results suggest that formant dynamics are likely informative for these contrasts, and that all three cues contribute to distinguishing between these vowels also when formant dynamics are considered (Figure 10).[10] Figures 9 and 10 further highlight the presumable diphthongization in [eː] and [oː], as indicated by the steepness of the fitted curve.

The GAMMs fit to the lower-mid front long and short vowel contrasts suggested an effect of vowel on F1 and F2 (all $ps < .0001$), but not on F3 (for [æː], $\hat{\beta} = -.012, SE = .01, p > .25$; for [æ], $\hat{\beta} = .012, SE = .008, p > .14$). Figure 11 demonstrates how these vowels overlap in F3-dynamics, but are distinguished for larger parts of the segment along F1 and F2.

Finally, for the GAMMs fitted to the mid center vowels, there was an effect of category on all cue evaluations for both long and short vowels, with the exception of the long vowels fit to F1 (for [øː] - [œː], $\hat{\beta} = -.017, SE = .01, p > .12$). For several of the other evaluations, the effect was relatively small, however significant (all $ps < .044$; Figure 12). These results would seem to suggest that when formant dynamics are considered, the [øː] - [œː] contrast is primarily upheld by F2 and F3.

The sets of neighboring contrasts investigated here all exhibited varying degrees of category overlap in static analysis. However, when formant dynamics was considered, the vowels in each contrast were all significantly different from each other along at least two cues (c.f., Fant, 1971; Kuronen, 2000; Pelzer & Boersma, 2019). For all contrasts, the vowels overlapped in parts of the segment, but importantly never along the entire segment (Figures 9**A-E**, 10**A-D-E**, 11**A**, 12**A-B-E-F**). This indicates that category overlap found in static analysis is mitigated once temporal analysis is included, which suggests that category distinctions unfold over time.

[10] An effect of gender was found for the high back vowels predicting F1 for the long and short vowels (both $ps < .022$), which suggests that the normalization approach likely reduced some talker-specific differences related to anatomical differences, but not all.

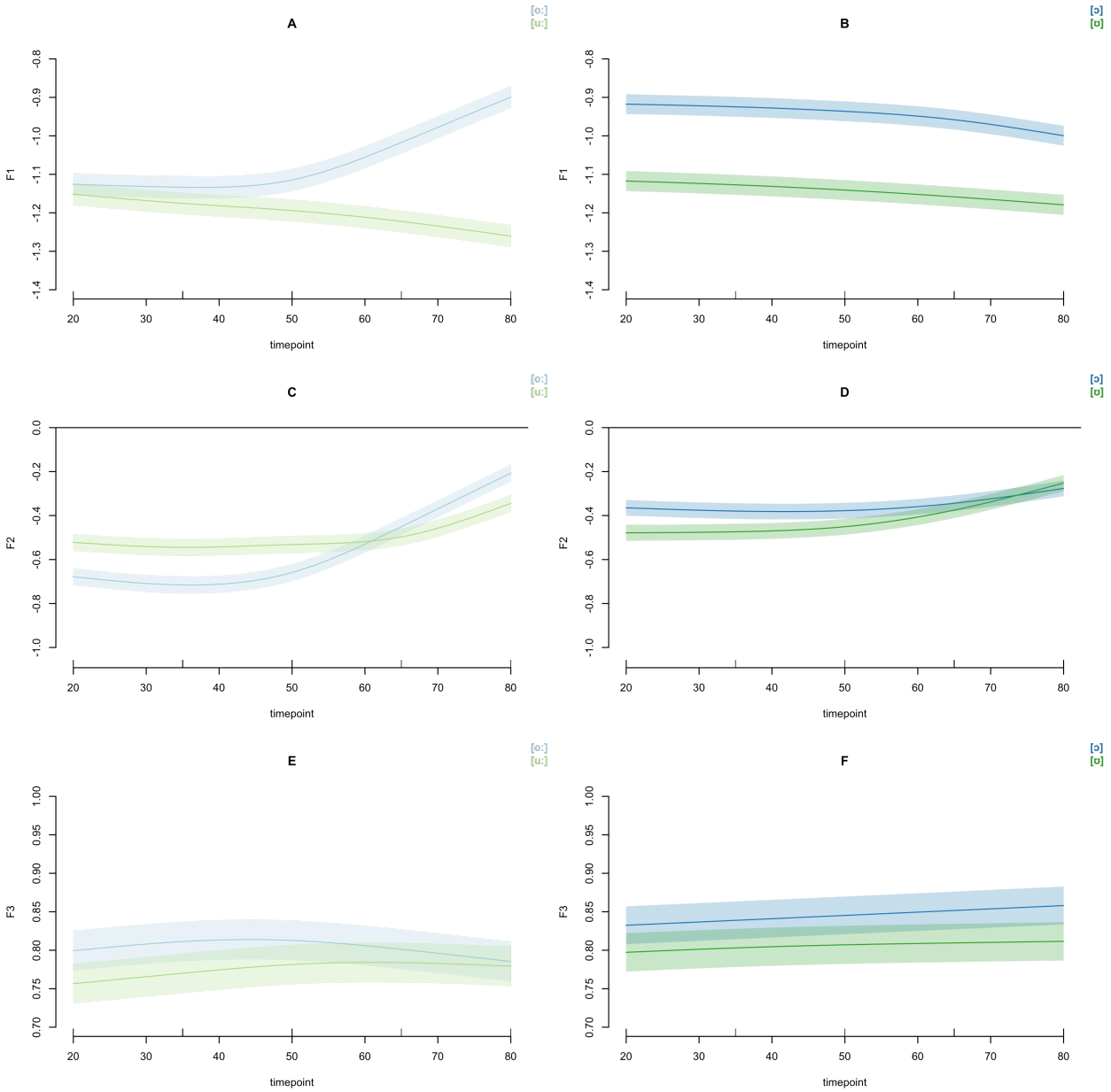

Figure 10. Fitted smooths of GAMM for predicting F1 (**upper row**), F2 (**mid row**), F3 (**bottom row**) and 95% confidence intervals for the long vowels [oː] - [uː] (**left**), and the short vowels [ɔ] - [ʊ] (**right**).

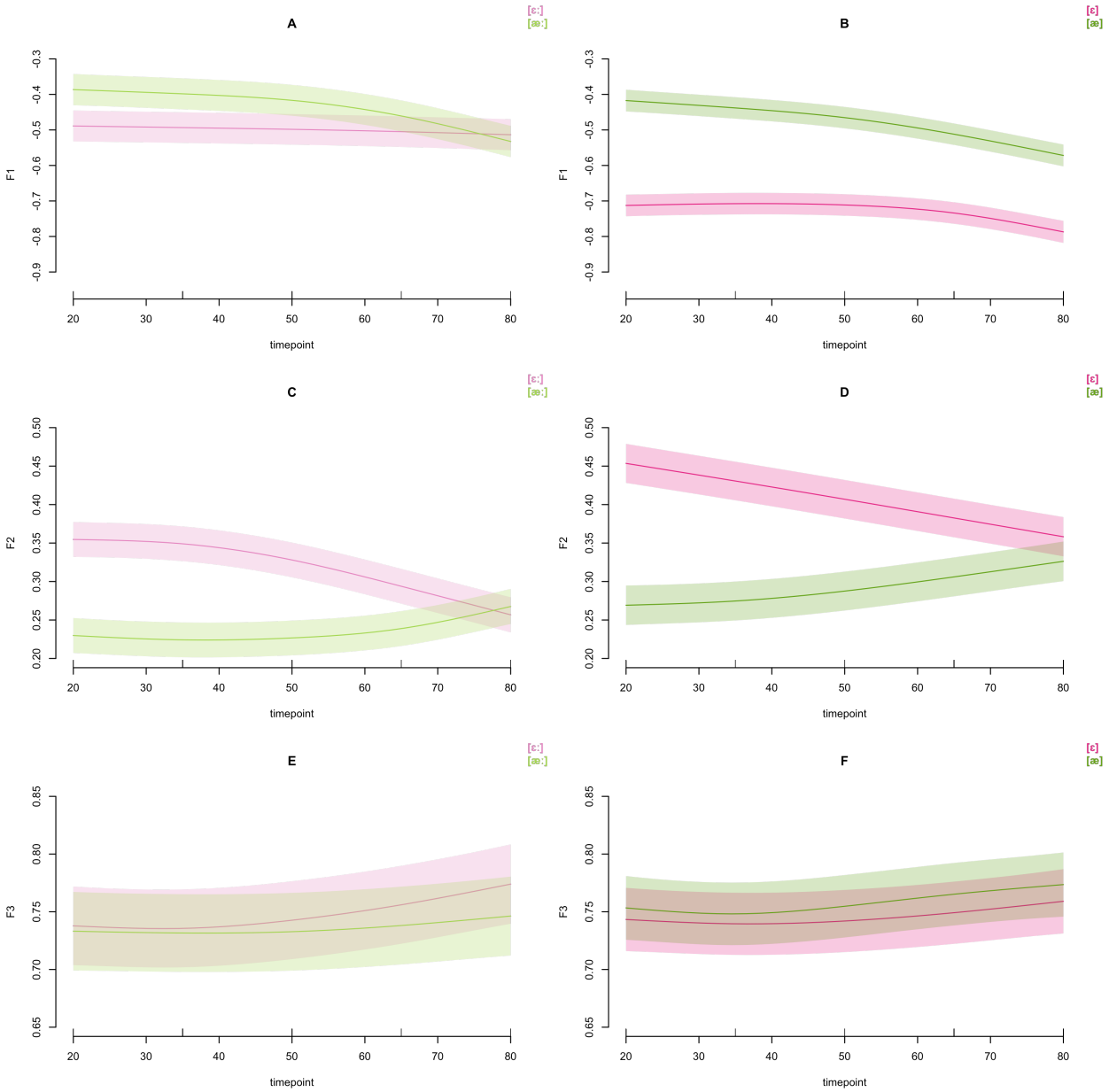

Figure 11. Fitted smooths of GAMM for predicting F1 (**upper row**), F2 (**mid row**), F3 (**bottom row**) and 95% confidence intervals for the long vowels [ɛː] - [æː] (**left**), and the short vowels [ɛ] - [æ] (**right**).

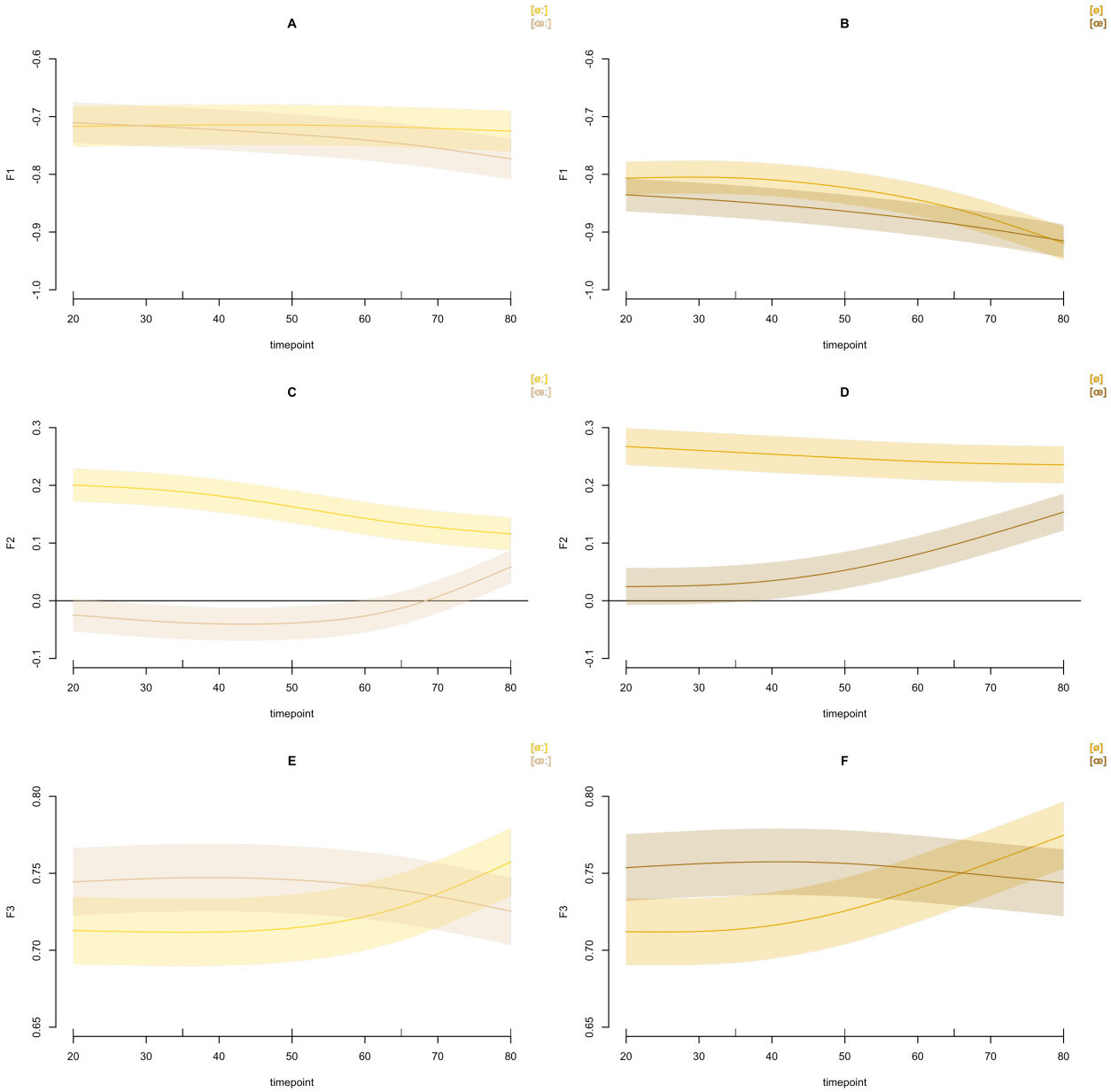

Figure 12. Fitted smooths of GAMM for predicting F1 (**upper row**), F2 (**mid row**), F3 (**bottom row**) and 95% confidence intervals for the long vowels [øː] - [œː] (**left**), and the short vowels [ø] - [œ] (**right**).

549 The results further indicate that F3-dynamics carry little information for the [iː] - [yː]
550 contrast. For the short [ɪ] - [ʏ], however, it is F1 rather than F3 that does not distinguish,
551 suggesting overlap in F1 (height), but not F3 (lip-rounding).

552 **2.2.2.2.2 Formant dynamics in long-short vowel pairs** The second set of
553 GAMMs modeled the effect of F1, F2, F3 and duration on all long and short vowel pairs.
554 Summary tables and visualizations of these GAMMs are included in the SI (Section S1.2.7).
555 There was a treatment effect of vowel on all spectral and temporal cues for all vowel pairs
556 (for F1, all $ps < .04$; for F2 and duration, all $ps < .0001$, for F3, all $ps < .0005$). With the
557 exception of the [yː] - [ʏ] vowel pair, all rounded vowels displayed lower F3 values, hence
558 more lip-rounding, in their long allophones (e.g., Hadding, Hirose, & Harris, 1976;
559 Stålhammar, Karlsson, & Fant, 1973). The results suggest that when formant dynamics
560 are considered, *all* long-short Central Swedish vowel pairs differ in spectral and temporal
561 cues. Among the pairs that displayed smaller, albeit statistically significant, differences in
562 spectral cues are the [iː] - [ɪ] pair predicting F1 and F3, possibly indicating a tendency for
563 stronger duration dependency, in line with previous perceptual work (Behne et al., 1997).
564 While the static analysis suggested that F1-F2-F3 only numerically increased separability
565 for long-short pairs, with the exception of [yː] - [ʏ] and [eː] - [ɛ], the dynamic analysis
566 suggests that all cues, when considered separately, carry information about vowel quality in
567 quantity contrasts under the assumption of formant dynamics.

568 ### 2.2.3 Results summary

569 The results from the static and dynamic analyses in Study 1 suggest that F1, F2, F3 and
570 duration all contribute to the distinction of Central Swedish vowels. While F1 and F2 are
571 or primary importance for most contrasts, F3 does contribute to increasing the separability
572 between some neighboring vowels differing in lip-rounding. The static analysis furthermore
573 suggested that the inclusion of F3 had less importance for the long-short vowel contrasts,

as statistically indistinguishable performance was found for the F1-F2 model.

Some vowels displayed overlap in the static analysis but increased separability when formant dynamics was considered, as indicated by formant trajectory analysis and GAMMs. The dynamic analyses highlighted that the short vowels also display formant movements, and that for most of both long and short categories, a larger portion of the dynamics resides in the later part of the segment. Given the increased separability of neighboring contrasts found in dynamic analysis, it is reasonable to assume formant movements as an auxiliary cue to vowel identity, more so for some contrasts than others. For instance, the [iː] - [yː] - [ʉː] - [eː], [ɪ] - [ʏ], [oː] - [uː], [ɔ] - [ʊ], [ɛː] - [æː], [ɛ] - [æ], [øː] - [œː], and [ø] - [œ] contrasts displayed considerable overlap in static analyses but increased distinguishibility when analysed dynamically.

While the static analysis suggested increased separability for the [iː] - [yː] contrast when F3 was included, this was not found under the assumption of formant dynamics. The GAMM fit to [iː] - [yː] - [ʉː] - [eː] contrast found no effect of [yː] relative to [iː] predicting F3. However, the GAMM fit to [yː] - [ʏ] suggested statistically significant differences between the two categories predicting F3. These two analyses taken together suggest more effects on F3 in the short vowel compared to the long vowel.

Table 2
The phonetic characterization of long (left) and short (right) Central Swedish vowels (as represented in the SwehVd database). Rounded vowels are shaded.

	front	central			back
high		[iː]	[yː]	[ʉː]	[uː]
mid-high	[eː]				[oː]
lower-mid	[æː]		[øː]	[œː]	[ɑː]

	front		central		back
high	[ɪ]	[ʏ]	[ɵ]		[ʊ]
mid-high	[ɛ]		[ø]	[œ]	[ɔ]
lower-mid	[æ]		[a]		

The resulting phonetic characteristics of the long and short Central Swedish vowels presented in Study 1, is summarized in Table 2. Beginning with the long vowels, there are 4 high vowels. The current acoustic description suggests that none of them are front. Instead, [iː] and [yː] group with [ʉː] as central vowels, and [uː] is back (c.f., Pelzer &

Boersma, 2019; Schötz et al., 2011).[11] There are 2 mid-high vowels—[eː] (front) and [oː] (back)—and 4 lower-mid vowels, [æː] (front), [øː], [œː] (both central) and [ɑː] (back). Given the substantial lowering of [ɛː] and its overlap with [æ], it is reasonable to assume one long allophone for /ɛ/, which is [æː] (c.f., Pelzer & Boersma, 2019).

The short vowel space contains 4 high vowels, two of which are front, [ɪ] and [ʏ], one is central [ʉ] and one back [ʊ]. There are 4 mid-high vowels, [ɛ] (front), [ø], [œ] (both central), and [ɔ] (back), and 2 low vowels, [æ] (front), and [a] (central). The analysis of this database supports what Riad (2014) anticipated and Pelzer and Boersma (2019) suggested, namely, a vowel system consisting of three height levels only, in contrast to the traditional four height levels system (e.g., Engstrand, 1999, 2004; Riad, 2014).

The motivation of the summarized acoustics presented in Table 2 rests on a pairwise grouping of the long and short vowels, similar to phonological analyses of Central Swedish (Riad, 2014). For instance, despite its high position, [eː] is defined as mid-high, on par with [oː], as both vowels share diphthongizational patterns, and their short versions are both lower than their long counterparts. Furthermore, [æː] is front rather than central as its short version is clearly more front than [ø]. Because of their overall centralized positions in SwehVd, [iː] and [yː] groups with [ʉː], while their short versions are still clearly front. One could thus argue that both /i/ and /y/ are under-specified for the front-back dimension, which would also be the case for /ɑ/, as [ɑː] is back and [a] is central. It is important to note that while Table 2 may point to possible updates of Central Swedish vowel phonology, this is only tentative as more evidence is required for a definite update. These include, e.g., more investigations in different contexts.

The phonetic characterization in Table 2 suggests that some categories have shifted in comparison to previous work. In order to assess the scope and spread of these changes, Study 2 was conducted.

---

[11] Whether [yː] can still be considered rounded given the high F3 values, is a question for future research.

3 Study 2: Vowel category movements over time - from SweDia to SwehVd

The aim of Study 2 was to investigate possible vowel shifts over the last generation by comparing the acoustic characteristics of SwehVd against that of 8 reference talkers in the SweDia materials, along the two primary cues to vowel identity, F1 and F2. Next, the reference materials used for comparison is presented, alongside the distance metric used for assessing potential vowel shifts.

3.1 Methods

3.1.1 Materials

The materials used for comparison is a database of N = 8 L1 talkers (N = 4 female) of Central Swedish that were recorded as reference talkers of Standard Swedish for the SweDia dialect database (Eriksson, 2004). All talkers were L1 talkers of Swedish, born and raised in the Greater Stockholm area or surroundings (provinces of Södermanland, Uppland), and of 22-35 years of age (mean age = 27; SD = 4.41). The talkers were recorded reading a list of 41 words containing all 21 vowels of Central Swedish (for the complete wordlist used for recording, see Table S33 in SI). Thirty-nine words on the list were monosyllabic, and two were bi-syllabic (*leta*, *flytta*). The target vowels extracted ([eː], [ʏ]) from the bisyllabic words had primary stress and were thus included in this study.

All [æ] vowels were excluded from the subsetted materials used here, as this vowel was recorded by only two of the talkers. All repetitions recorded by a talker were included, even though the variability in the number of repetitions for each vowel across talkers was substantial.[12] The phonological contexts used were the same across talkers (with two exceptions, see Table S33 in SI) but differed across categories. Importantly, none of the

[12] Three to 8 repetitions of [iː], [yː], [ʉː], [ɵ], [eː], [æː], [ø], [œː], [œ], [ɑː], [a], [oː], [uː], [ʊ], 6 repetitions of [ɪ], [ʏ], [ʏ], [ɛ], 3 to 14 repetitions of [øː], [ɔ], [ɛː] per talker.

643 contexts were hVd. For details on the recruitment, recording, pre-processing and
644 annotation procedure in the SweDia materials, see Eriksson (2004).

645 The two databases were of unequal sample sizes overall: N = 9103 datapoints for
646 SwehVd, and N = 669 datapoints for SweDia. The regional origin of the talkers matched
647 across databases, age at the time of recording (mean age in SweDia = 27, mean age in
648 SwehVd = 29) and the relative proportion of male and female talkers were also highly
649 similar (50% female talkers in SweDia2000, 55% female talkers in SwehVd). Any potential
650 effect of differences in gender distributions on the comparison were presumably reduced by
651 the use of normalized vowel data.

652 Cue measurements from SweDia were extracted in the same way as for SwehVd
653 (Section 2.1.2.1). To correct for measurement errors in the automatic extraction of cues, 5
654 separate univariate distributions of the five extracted cues (F0, F1, F2, F3 and duration)
655 was estimated for each distinct combination of talker and vowel. Points that fell outside of
656 the $0.25th$ to the $97.5th$ quantile of the distributions for each vowel were identified,
657 examined for measurement errors, and subsequently corrected. This followed the approach
658 employed for the SwehVd corpus (Persson & Jaeger, 2023), and strikes a middle-ground
659 between the ideal (manual correction of all tokens) and feasibility. Outliers were identified
660 and removed following the approach for the subsetted SwehVd dataset used in Study 1,
661 leaving N = 664 datapoints for analysis.

662 3.1.2 Assessing vowel shifts

663 In order to assess whether the relative placement of the categories in the acoustic space was
664 different between datasets, formant data was visualized in F1-F2 space and distances
665 assessed with the *orthogonal projection ratio*, henceforth *op* (M. Stevens, Harrington, &
666 Schiel, 2019).[13] Both visualizations and *op* calculations were based on measurements at the

[13] Other distance measures are possible, such as Mahalanobi's distance. The primary advantages of using the *op* is that it considers the overall structure of vowel spaces rather than individual distances, and that it

steady-state portion of the vowel in both datasets—taking the average F1-F2 values across the three mid-points (at 35, 50, and 65% into the vowel). This approach presumably reduced the risk of coarticulatory effects resulting from the use of different phonological contexts in the datasets.

3.1.2.1 Orthogonal projection ratio The *op*, formalized in (2), compares the position of a single token of a V_{vowel} from a talker, relative to the means of two anchor categories, U_{anchor} and W_{anchor}, from the same talker. The anchor categories are selected based on their position in the space relatively to V_{vowel} (for an illustration, see SI Figure **??**). After calculating the *op* for each token of V_{vowel} from each talker relative to the anchors, the *op* is summarized across all tokens for that talker. This procedure is repeated for each of the vowels and their respective anchors for each talker in each dataset and then compared across datasets, in order to assess vowel shifts over time.

$$op(V_{vowel}) = (-2 \cdot \frac{((V_{vowel} - U_{anchor})\%*\%(U_{anchor} - W_{anchor}))}{((U_{anchor} - W_{anchor})\%*\%(U_{anchor} - W_{anchor}))}) - 1 \quad (2)$$

An $op = 0$ indicates that the vowel V_{vowel} is positioned exactly in the middle between U_{anchor} and W_{anchor}, an $op = \pm 1$ indicates that V_{vowel} is positioned exactly on the mean of one of the two anchors, whereas a V_{vowel} positioned somewhere between U_{anchor} and W_{anchor} has an *op* that varies between ± 1. The *op* can thus be used as an index of e.g., lowering, fronting, or centralizing.

Table 3 lists all the vowels V_{vowel} tested along with their corresponding anchor points, U_{anchor} and W_{anchor}. These were all vowels for which the overall distance in F1 and F2 vowel means in SwehVd to SweDia, $\geq .85$ standard deviations from the vowel means in

offers a standardized way to compare the degree of category membership across different datasets.

SweDia.[14]

Table 3
The anchors used for calculating orthogonal projection ratio across the two datasets

U_{anchor}	V_{vowel}	W_{anchor}
[ɑː]	[ʏ]	[ɪ]
[ɪ]	[ʉː]	[oː]
[eː]	[ɛː]	[æː]
[eː]	[øː]	[ɑː]
[ʉː]	[ø]	[a]
[ɵ]	[œ]	[œː]
[ɑː]	[ɔ]	[oː]

To assess the magnitude of differences in *op* between datasets, separate LMMs for each of the categories investigated were fit, predicting *op* from database, with talkers as random effect: $op \sim Database + (1|Talker)$. Database was treatment coded with SweDia as reference level.[15]

3.2 Results

Figure 13 displays the vowel movements in F1-F2 space across talkers over time. The figure suggests that a few vowels display considerable stability over time, e.g., the back vowels [ɑː] and [oː], the center vowels [œː] and [a], and the front vowels [eː] and [ɛ]. The remaining vowels have all moved to a lesser or greater extent. Two general tendencies can be noted. Firstly, several of the vowels that form the edges of the space slightly shift further outwards, causing the overall space to widen. This is the case for [ɪ], [æː], [oː], and [uː]. Most of the inner vowels either also shift outwards—[ʏ], [ɵ], [ʉː], [ʊ], [ɔ]—or centralize—[ø],

[14] The front [æː] was also included in this subset. Given that it occupies the lowest position in the space, the calculation of *op* was not possible.

[15] The two databases employed different phonological context. One way to control for this difference would be to use random effects by word. However, this was not possible since this random effect would be highly collinear with the database, as SwehVd only used one word context. Indeed, most effects were no longer significant when word was included. Furthermore, it is reasonable to assume that coarticulatory effects from using different phonological contexts would be minimal given that the measurements were extracted from the steady-state of the vowel. Therefore, excluding word as random effect likely had very little effect on the results.

[øː], [œ]. Second, these movements lead to some neighboring categories closing in on each other, e.g., [ɪ] - [ʏ], [iː] - [ʉː], [ø] - [œ], [oː] - [ɔ] - [uː] - [ʊ], or increasing their separability—[ɵ] - [œ], [ø] - [øː], [ʏ] - [ʉː].

Figure 13 further suggests a series of consecutive movements in the front part of the space, where [eː], [ɪ], [ʏ] have traveled further up and front in SwehVd, while [iː], [yː] and [ʉː] have centralized. Finally, Figure 13 shows that [ɛː] have lowered substantially compared to SweDia, down to a position close to [æː] (and to [æ]; recall Figure 1). The already low [æː] has in turn moved further down, presumably constrained by the lower bound and cannot move further.

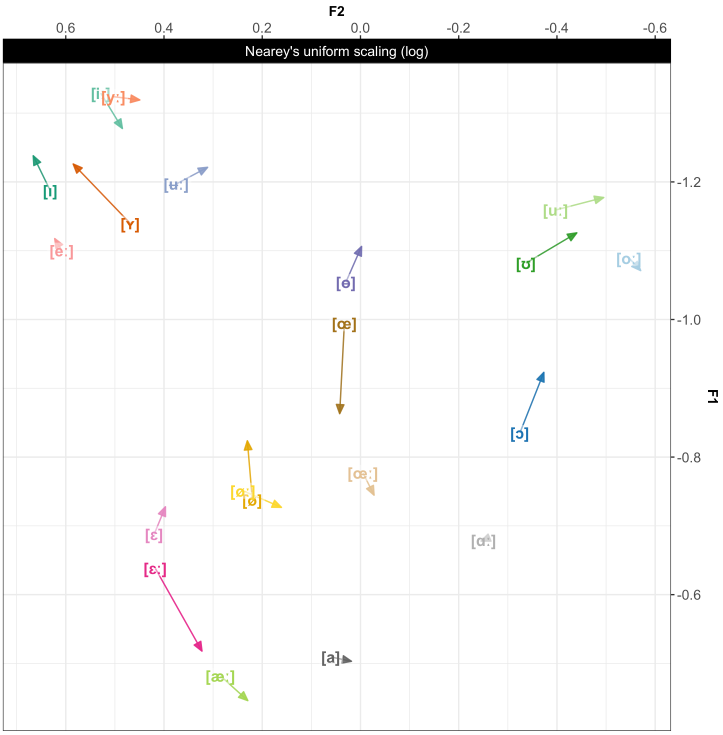

Figure 13. Vowel category movements in the Central Swedish vowel space from 1999 to 2023, in F1-F2 space. The arrow indicates the direction of movement over time, with base of arrow representing SweDia (vowel label) and tip of arrow representing SwehVd. Vowel data is taken at the steady state of the vowel (averaged across the three mid-points) and averaged across talkers in the two datasets.

To quantify these potential vowel shifts, the distances between the category locations in the two datasets were assessed with the *op*. Figure 14 visualizes the *op* of all vowels in

Table 3 in both datasets. By this way of assessing the relative movement between anchor points across datasets, Figure 14 indicates that all selected categories have moved over time in the directions suggested by the qualitative analysis in Figure 13. The largest differences appear to concern [ʏ], [ɛː], [œ], and [ɔ]. The front [ʏ] is relatively closer to [ɪ] in SwehVd than in SweDia, which suggest fronting. The [ɛː] is relatively closer to [æː] in SwehVd than in SweDia, as is [œ] to [œː], indicating the lowering of both categories. The back [ɔ] is relatively closer to [oː] in SwehVd, indicating back raising.

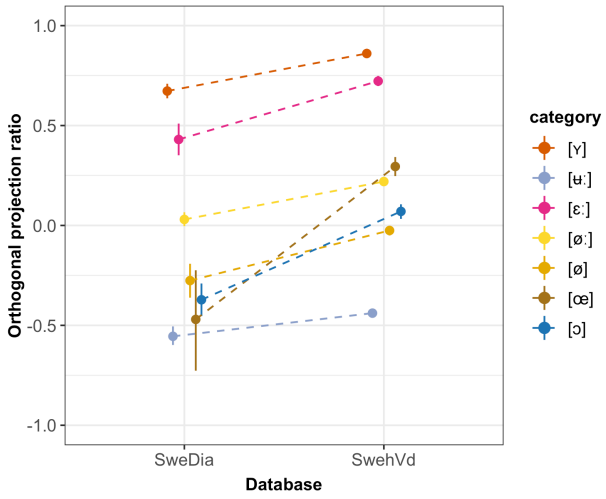

Figure 14. The relative distance between categories in SweDia and SwehVd. $op = 0$ indicates that the vowel is equidistant between the two anchor categories in Table 3, positive numbers indicate a position closer to the second anchor vowel, negative numbers indicate a position closer to the first anchor. Pointranges indicate mean and 95% bootstrapped CIs of the *op* summarized across talkers in each database. The large CIs for the [œ] vowel in the SweDia materials likely reflect the tendency of some talkers to neutralize the short /ø/ and /ʉ/ as [ɵ].

The LMMs fit to the data confirmed the directionality of the vowel shifts, and found a main effect of Database on *op* for all categories (all $ps < .015$), with the exception of [ʉː] ($\hat{\beta} = 0.11, SE = 0.06, p > .07$). Summary tables of all 6 models are included in the SI, Section S1.3.4.

A discussion of these results, and the results from Study 1, follows next.

4 General discussion

The objective of the present paper was to provide an extensive acoustic analysis of contemporary Central Swedish vowels and to evaluate the extent of vowel shifts over the last generation. The main findings from both studies are next discussed, alongside methodological considerations and future directions.

4.1 Static and dynamic analyses of the 21 vowels of modern-day Central Swedish

The purpose of the first study was to present up-to-date static and dynamic acoustic analyses of Central Swedish vowels that included both empirical formant data and models of formant dynamics. The first study thus aimed to expand on and complement previous work by 1) the scope of the analysis, performing the same type of analyses on all 21 categories, 2) the materials chosen, using a recently collected hVd corpus with high resolution within and across talkers for a single variety, and 3) the methodological approach employed, with the use of traditional formant analysis, category separability index, trajectory visualizations and models of formant dynamics (GAMMs). Study 1 furthermore aimed to evaluate the hypothesized importance of F3 for contrasting rounded vs. unrounded categories, the extent to which *all* long-short vowel pairs display spectral differences, and what part of the vowel space is more susceptible to diphthongization.

Beginning with the static analysis, the results seem to suggest that the most important cues to vowel identity were F1, F2 and duration, which replicated previous work (e.g., Kuronen, 2000; Lindblom, 1963). For some contrasts, F3 increased separability, highlighting the F3 dependency for certain rounding contrasts. The category separability index calculated for the long-short vowel pairs suggested that categories in each quantity contrasts were, with the exception of /y/ and /e/, just as separable in F1-F2 as F1-F2-F3 space. This would suggest differences in lip-rounding for the long and short /y/, that were

not found for the other rounded vowels. The results furthermore suggested that static measurements across talkers, while informative, were insufficient to accurately capture the spectral acoustics of some vowels, as indicated by the amount of overlap when static formants were considered. A more exhaustive description can be achieved by including dynamic analyses, as shown in this paper.

This was especially true for the [ɛː] - [æː], [uː] - [oː] and [eː] - [iː] - [yː] distinctions, given the direction and scope of formant trajectory movements across the vowel segment. The analysis of the empirical data furthermore suggested that a larger portion of the movement took place at the later part of the segment for most vowels, irrespective of the magnitude of change in F1 and F2. However, [iː], [yː] and [ʉː] constituted important exceptions as they displayed more movement in the first three to four time-points. Of note, some of the short vowels showed formant movements of equal or larger magnitude as certain long vowels, which seems to signal the importance of vowel dynamics for long and short vowels alike. This has been investigated in work on other languages (e.g., Hillenbrand et al., 1995; Watson & Harrington, 1999), but has largely been lacking in studies on Swedish. In terms of distinguishing between diphthongization and merely formant movement, the trajectory plots suggested diphthongization towards an open quality in primarily [eː] and [oː].

GAMMs fit to the data contributed with further insights into the formant dynamics of individual contrasts as well as for vowels differing in quantity, and allowed for an assessment of the relative contribution of formant dynamics to cue dependencies for neighboring and more distant contrasts. For instance, GAMMs fitted to the long-short vowel pairs indicated that all pairs were reliably distinguished by all included spectral cues, highlighting the informativity carried by formant dynamics in quantity distinctions. The static and dynamics analyses thus both suggest that quantity distinctions are not separate from quality distinctions in Central Swedish.

With regard to the neighboring contrasts hypothesized to rely on formant dynamics—the [iː] - [yː] - [ʉː] - [eː], [oː] - [uː], [ɛː] - [æː], and [øː]-[œː] contrasts, and their short

counterparts—the results indicated significant differences for all comparisons and all cues with the exception of [iː] - [yː], [ɛː] - [æː], and [ɛ] - [æ] predicting F3, and [ɪ] - [ʏ], [øː]-[œː] predicting F1. This would seem to suggest that the movements in these vowels along these cues, are not contributing to vowel quality information. Presumably, the dynamics in remaining cues is sufficient for distinguishibility.

Since the GAMMs were fit to each cue separately, they allowed for an assessment of the relative weight of each cue, compared to the separability index where the by-cue contributions can only be evaluated indirectly. An inherent limitation in the separability index, as implemented in this paper, is the simplifying assumption that all dimensions within a cue combination carry equal weight. It is therefore not possible to assess whether the relation between the cues in each space is symmetrical or not, that is, in the F1-F2 space, we do not know whether F1 carries as much information as F2 for separability, and vice versa. In addition, given that the comparisons are pairwise, they are limited to explaining the relation between two vowels in a contrast. As such, they cannot inform us of the separability of a given vowel from other neighboring vowels, or the overall category separability in the entire space. This is, however, a limitation that the separability index shares with the GAMMs. Neither of these methods are able to assess the distinguishability or confusability of all vowels under different cue combinations in one analysis. Nor can they inform us of the *perceived* distinguishibility, even if it is reasonable to assume that reduced overlap between tokens of neighboring categories would increase intelligibility (e.g., Bradlow, 1995; Wright, Local, Ogden, & Temple, 2004). For this, one could either fit a multinomial model predicting vowel from cue combinations, or a perceptual model that can assess the predicted consequences for perception by the use of categorization accuracies and confusion matrices. Such models can be an important avenue for future research, a starting point for the design of perception studies that can shed more light on the consequences of the present results for the perception of Central Swedish vowels. For instance, in a language with a systematic quantity distinction such as Swedish, the role of spectral cues

in long-short vowel pair distinctions could be assessed by exposing listeners to synthesized versions where long and short vowel duration is crossed with the allophones' spectral information for any given phoneme. Furthermore, as the results seem to support claims of the hypothesized importance of formant dynamics for vowel distinctions, more insight into the effect of formant dynamics for vowel perception could be gained from having listeners categorize tokens extracted from different segments of the long vowels, e.g., the first three time-points vs. the three final time-points (c.f., Jenkins, Strange, & Miranda, 1994; Strange, 1989). The design of such experiments can be informed by modeling the predicted perceptual consequences of different cue spaces, and of considering different vowel segments.

4.2 Vowel shifts

In Study 2, the acoustic characteristics of SwehVd was subsequently compared to that of 8 reference talkers of Central Swedish from the SweDia materials recorded approximately one generation ago, in order to investigate the scope and spread of vowel changes noted in Study 1 (some of which, previously reported in Pelzer & Boersma, 2019). The aim was to present a broad-scale comparison by evaluating changes in all but one of the 21 categories, using empirical formant analysis and the *op*.

Several vowels had shifted quite considerably compared to earlier work on Central Swedish (c.f., Fant et al., 1969; Kuronen, 2000). Two of the most important shifts concerned the fronting of [ɪ] and [ʏ] and the centralization of [iː] and [yː], where [ɪ] and [ʏ] appeared to maintain their positions as high front vowels, while [iː] and [yː] had centralized to the mid-center part of the space. Compared to the SweDia reference materials, the shifts in [ɪ], [iː] and [yː] were not as substantial as for other vowels, which suggests that these changes had begun already in 1999, and then continued throughout the first two decades of the 2000s (see also, Pelzer & Boersma, 2019).[16]

[16] Unfortunately, Pelzer and Boersma (2019)'s study on diphthongization only included the long vowels, it is therefore difficult to know whether the fronting of the short vowels was as pronounced in 2019.

The high front vowels form a particularly interesting part of the space given that several possibly related processes might be ongoing. The visualizations and GAMMs suggested that F3 dynamics did not contribute to the distinction between [iː] and [yː], and that [yː] appeared to be less rounded than [ʏ]. This finding conflicted with previous work on Central Swedish (e.g., Fant et al., 1969; Fujimura, 1967; Kuronen, 2000), and would seem to indicate that the [iː] - [yː] contrast might primarily be supported by F1-F2 dynamics, or by additional acoustic cues not investigated in this study. There is of course also the possibility that listeners might disambiguate the two categories using primarily visual cues or linguistic information, or perhaps these two categories are not distinguished, which would suggest a merger in process among these talkers. A merger might be driven by relaxation of lip-rounding, as supported by the higher F3 values found for [yː].

Both of these processes—centralization and relatively sustained overlap—could however be explained by [iː] and [yː] being produced as damped versions. The presence of a damped [ɨː] would be supported by the lower F2 values, as the consonantal offglide in [ɨː] lowers F2 (Engstrand et al., 2000). A merger of [iː] and [yː] into [ɨː] has been observed among younger talkers in other regions in Sweden, e.g., for Gothenburg Swedish, as reported by Gross and Forsberg (2020). If centralization of [iː] is a prerequisite for such a merger, as suggested by Gross and Forsberg (2020), the present results might indicate the beginning of a merger. A future scenario might involve the loss of the defining feature lip-rounding in [yː] if a merger continues, given its weaker perceptual salience (as hypothesized by Gross & Forsberg, 2020).[17]

Impressionistic listening by the author did support the presence of a buzzing sound, similar to [ɨː], among the majority of talkers in SwehVd, both male and female. The

[17] A parallel exist in iotacism in Norwegian and Swedish dialects (Eliasson, 2000), characterized by the loss of lip-rounding resulting in vowel shifts such as [yː] to [iː] for monophthongs, and [øy] to [ei] for diphthongs. Iotacism describes a feature loss for a fairly unusual phonological characteristic—rounded front vowels. Talkers that merge because of iotacism would not necessarily produce [iː] as the damped [ɨː], which could potentially explain the vowel patterns of some talkers in SwehVd that merge the two vowels but have a rather fronted [iː] without an offglide consonant element.

strength and scope of [iː] varied across talkers, from a relatively strong [iː] (17 talkers), to a weaker buzzing sound (12 talkers), or more of a consonant offglide element similar to [j] following [iː] (3 talkers). Interestingly, several of the talkers that did not produce any apparent final consonant glide or buzz (11 talkers), seemed to have overall less retracted [iː] and [yː], hence supporting the hypothesized link between centralization and consonantal offglide. Further insights into these individual differences can be gained by studying the SwehVd materials on a talker-specific level.

The changed placement of [iː] and [yː] might have affected the surrounding front vowels. For instance, [eː] was *not* the most fronted vowel for some of the talkers who did not centralize [iː] and [yː], which would seem to suggest that the fronted placement of [eː] might be conditioned by a more centralized position of [iː] and [yː]. The order and cause of these events is of course unclear, however, they might suggest underlying structural co-variation, such as chain-shifts, causing several categories to move (e.g., Brand, Hay, Clark, Watson, & Sóskuthy, 2021). Chain-shifts are hypothesized to be causally related, in that the shift in one vowel sets off changes in other vowels (Gordon, 2013; Hay, Pierrehumbert, Walker, & LaShell, 2015; Maclagan & Hay, 2007; Martinet, 1952). Push-chain-shifts are characterized by a moving vowel encroaching on another vowel's space, causing that vowel to move, whereas pull-chain-shifts describe a situation where a travelling vowel leaves a vacuum for another vowel to enter (Gordon, 2013; Łubowicz, 2011). Underlyingly, this structural co-variation might be driven by system-wide evolutionary pressure, i.e. the drive towards symmetry (Boersma, 1998) or vowel spaces that maintain or maximize contrast or dispersion (Liljencrants & Lindblom, 1972; Martinet, 1952; Schwartz, Boë, Vallée, & Abry, 1997). The centralization of [iː] and [yː] could thus have pushed [ɪ] and [ʏ] to travel front in order to maintain a contrast between the long-short pairs. Alternatively, the fronting of [eː] might have pushed back [iː] and [yː], however, what might be causing [eː] to move remains unclear, given that the neighbouring vowel [ɛː] substantially lowers. A future scenario might involve [eː] once more descending to a position closer to its shorter counterpart, and

closer to the position it occupies in the latter part of the segment (see Figure 7). Such a change would potentially result in a sparser front space, possibly setting the stage for a more compact system overall. Future studies should aim to assess this co-variation in a principled way for further insight into which of these factors might be driving the shifts.

Other noticeable changes outside of the high front area concern [ɛː] and [œ], that have both lowered since 1999. The lowering of [ɛː] was anticipated by e.g., Riad (2014), and confirmed in T. Leinonen (2010) and Pelzer and Boersma (2019). Even if [œ] has lowered compared to SweDia, [øː] and [ø] are still located beneath [œ], which suggest a backing rather than lowering of /ø/ in position before retroflex segments (c.f., Riad, 2014). Furthermore, the difference in placement of [œ] might result from a presence (SweDia) or lack of (SwehVd) neutralization of [ø] into [ɵ]. In both materials, [ø] precedes /r/ (SwehVd - *hörr*, SweDia - *dörr*), and in SweDia, [ø] is considerably closer to [ɵ] than in the SwehVd. If this result holds, it would suggest that the previously documented merger of short [ø] and [ɵ] might have lost its importance in Central Swedish over the last generation (Ståhle, 1965; Wenner, 2010; but see Kotsinas, 1995 for younger Stockholm talkers).

4.3 Methodological considerations

Methodological considerations for the present study that goes beyond those already discussed concern 1) the choice of how to measure some of the acoustic cues, 2) methods for assessing ongoing changes in the vowel space, and 3) materials used for comparing vowel changes over time. The first consideration concerns two of the cue measurements, F0 and duration, and how they were measured and evaluated. Here, the average F0 across the vowel segment was reported. However, pitch contours have been known to influence perceived duration and prominence of vowels (e.g., Gussenhoven & Zhou, 2013). One could therefore claim that a more accurate acoustic representation of vowels should have included pitch contours. Similarly, the temporal analysis of the effect of duration on long-short vowel pairs could be supplemented with measures of consonant ratios (e.g., Pelzer &

Boersma, 2019; Schaeffler, 2005). These limitations can be addressed in future studies as the SwehVd database is publicly available in an online repository (https://osf.io/ruxnb/).

Second, many acoustic-phonetic studies often include systematic listening by trained phoneticians, validated through measures of inter-rater reliability (for a review, see Cucchiarini, 1995; Gross & Forsberg, 2020; Kuronen, 2000; Pelzer & Boersma, 2019). In this study, proposals of the presence of a possible ongoing merger and/or centralization of [ɪː] - [ʏː], and the absence of merger of [œ] - [ɵ], would all receive stronger support if supplemented by systematic (and not impressionistic) listening. Furthermore, the amount and scope of vowel changes were assessed using the *op*. This non-parametric measure calculates each vowel's movement between two anchor vowels separately, as such, it does not take the entire system into account and can therefore miss patterns of co-variation within and across talkers. Employing a modeling approach similar to what Brand et al. (2021) suggest, could provide a more principled way of assessing vowel change system-wide, while accounting for by-talker variability.

Third and finally, different phonological contexts were employed for recording of the two databases compared, which can constrain claims made about vowel changes given differences in potential effects of coarticulation. This limitation was partly mitigated by selecting the steady-state portion of the vowel for comparison. A more comparable database would of course strengthen claims of vowel shifts. Unfortunately, given that existing databases on Central Swedish display substantial variation in composition (Table 1), there would inevitably be some loss in comparability irrespective of what reference materials chosen. The SwehVd sought to address this issue by recording hVd words, with the intention of enhancing comparability in studies on Central Swedish, and with the welcome side-effect of facilitating cross-linguistic investigations.

5 Conclusions

The present study has reported on the acoustic properties of Central Swedish vowels and how they have changed over the last generation. The spectral and temporal cues investigated all contributed to distinguishing between the 21 vowels in the Central Swedish vowel space, with varying weight. More insight into formant dynamics within and between quantities have been gained by the dynamic analysis presented, which is also of value for cross-linguistic research. What has been gained with the broad-scale approach of characterizing the *entire* vowel space adopted here, is of course lost in terms of detailed investigations of individual vowel contrasts. There is certainly a lot more to say about the centralization of [iː] - [yː], the potential loss of lip-rounding in [yː], the direction of vowel shifts, and lowering of [ɛː], among other things. The acoustic descriptions outlined in this paper, together with the publicly available SwehVd database, can provide a reference point for future investigations into these acoustic events and beyond.

Ethics statement

This study on human participants was granted an exemption from requiring ethics approval in accordance with the local legislation and institutional requirements (Etikprövningsmyndigheten, Uppsala, Sweden). The participants provided their written informed consent to participate in this study.

Data availability statement

The SwehVd dataset presented in this study can be found in an online repository (SwehVd: https://osf.io/ruxnb/). All analyses and visualization code can be found in a separate online repository (https://osf.io/7uvj4/).

Funding

The work presented in this study was partially funded by a grant from the Kinander's foundation (2021), a grant from Kungliga Vetenskapsakademien (2023), and by the Department of Swedish Language and Multilingualism at Stockholm University.

Acknowledgments

Omitted for review

Conflict of Interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

6 References

Assmann, P. F., & Katz, W. F. (2005). Synthesis fidelity and time-varying spectral change in vowels. *Journal of the Acoustical Society of America*, *117*(2), 886–895. https://doi.org/10.1121/1.1852549

Baayen, H., Vasishth, S., Kliegl, R., & Bates, D. (2017). The cave of shadows: Addressing the human factor with generalized additive mixed models. *Journal of Memory and Language*, *94*, 206–234.

Barreda, S. (2021). Perceptual validation of vowel normalization methods for variationist research. *Language Variation and Change*, *33*(1), 27–53. https://doi.org/10.1017/S0954394521000016

Barreda, S., & Nearey, T. M. (2018). A regression approach to vowel normalization for missing and unbalanced data. *The Journal of the Acoustical Society of America*, *144*(1), 500–520. https://doi.org/10.1121/1.5047742

Behne, D. M., Czigler, P. E., & Sullivan, K. P. H. (1997). Swedish Quantity and Quality: A Traditional Issue Revisited. *Reports from the Department of Phonetics, Umeå University*, *4*, 81–83.

Björsten, S., & Engstrand, O. (1999). Swedish "damped" /i/ and /y/: Experimental and typological observations. *Proc. 14th ICPhS*. San Fransisco.

Bleckert, L. (1987). *Centralsvensk diftongering som satsfonetiskt problem [Diphthongization in Central Swedish as a problem of sentence phonetics].* Uppsala: Skrifter Utgivna Av Institutionen För Nordiska Språk Vid Uppsala Universitet.

Boersma, P. (1998). *Functional phonology: Formalizing the interactions between articulatory and perceptual drives*. Amsterdam: Den HaagHolland Academic Graphics/IFOTT.

Boersma, P., & Weenink, D. (2022). *Praat: Doing phonetics by computer [Computer program].*

Bradlow, A. R. (1995). A comparative acoustic study of english and spanish vowels. *The Journal of the Acoustical Society of America*, *97*(3), 1916–1924.

Brand, J., Hay, J., Clark, L., Watson, K., & Sóskuthy, M. (2021). Systematic co-variation of monophthongs across speakers of new zealand english. *Journal of Phonetics*, *88*, 101096.

Bruce, G. (2009). Components of a prosodic typology of Swedish intonation. In *Components of a prosodic typology of Swedish intonation* (pp. 113–146). De Gruyter Mouton. https://doi.org/10.1515/9783110207569.113

Chuang, Y.-Y., Fon, J., Papakyritsis, I., & Baayen, H. (2021). Analyzing phonetic data with generalized additive mixed models. In *Manual of clinical phonetics* (pp. 108–138). Routledge.

Cucchiarini, C. (1995). Assessing transcription agreement: Methodological aspects. *Clinical Linguistics and Phonetics*, *10*(2), 131–155. https://doi.org/10.3109/02699209608985167

Eklund, I., & Traunmüller, H. (1997). Comparative Study of Male and Female Whispered and Phonated Versions of the Long Vowels of Swedish. *Phonetica*, *54*(1), 1–21. https://doi.org/10.1159/000262207

Elert, C.-C. (1964). *Phonologic studies of quantity in swedish: Based on material from Stockholm speakers*. Uppsala: Almqwist & Wiksell.

Elert, C.-C. (1980). Diftongeringar och konsonantinslag: Drag i uttalet av långa vokaler i svenskan av i dag [Diphthongizations and consonantal offglides: Traits in the pronunciation of long vowels in the Swedish of today]. *Språken i vårt Språk*, 168–181.

Elert, C.-C. (1981). *Ljud och ord i svenskan [Sounds and words in Swedish language]*. Umeå: Universitetet i Umeå, Almqvist & Wiksell international.

Elert, C.-C. (1994). Indelning och gränser inom området för den talade svenskan: En aktuell dialektografi [Distribution and boundaries within the area of spoken

Swedish: An up-to-date dialectography]. In *Diabas*: *Vol. 4. Kulturgränser - myt eller verklighet?* (pp. 215–228). Institutionen för nordiska språk vid Umeå Universitet.

Elert, C.-C. (2000). *Allmän och svensk fonetik [General and Swedish phonetics]* (8., omarb. uppl). Stockholm: Norstedt.

Eliasson, S. (2000). Typologiska och areallingvistiska aspekter på de nordeuropeiska språkens fonologi [Typological and area linguistic aspects of the phonology of the Northern European languages]. *In: Ernst Håkon Jahr (Ed.), Språkkontakt — Innverknaden Frå Nedertysk På Andre Nordeuropeiske Språk, 21–70. (Nord 2000:19.) København: Nordisk Ministerråd, 2000*, 21–70.

Eliasson, S. (2022). The phonological status of Swedish au and eu: Proposals, evidence, evaluation. *Nordic Journal of Linguistics*, 1–42. https://doi.org/10.1017/S0332586522000233

Engstrand, O. (1999). Swedish. In *Handbook of the International Phonetic Association: A guide to the usage of the International Phonetic Alphabet.* Cambridge: Cambridge University Press.

Engstrand, O. (2004). *Fonetikens grunder [The basics of phonetics].* Lund: Studentlitteratur.

Engstrand, O., Björsten, S., Lindblom, B., Bruce, G., & Eriksson, A. (2000). Hur udda är Viby-i? Experimentella och typologiska observationer [How peculiar is Viby-i? Experimental and typological observations]. *Folkmålsstudier*, *39*, 83–95.

Eriksson, A. (2004). SweDia 2000: A swedish dialect database. In P. J. Henrichsen (Ed.), *Babylonian confusion resolved. Proceedings of the nordic symposium on the comparison of languages, no. 1 in copenhagen working papers in LSP* (pp. 33–48).

Fant, G. (1959). Acoustic analysis and synthesis of speech with applications to swedish. *Ericsson Technics*, *15*, 3–108.

Fant, G. (1971). Notes on the Swedish Vowel System. In L. Hammerich, R. Jakobson, E. Zwirner, & E. Fischer-Jørgensen (Eds.), *Form and substance: Phonetic and linguistic papers.* Odense: Andelsbogtrykkeriet.

Fant, G. (1983). Feature analysis of Swedish vowels - a revisit. *STL-QPSR*, *24*(2-3), 001–019.

Fant, G., Henningsson, G., & Stålhammar, U. (1969). Formant frequencies of Swedish vowels. *STL-QPSR*, *10*(4), 026–031.

Fujimura, O. (1967). On the Second Spectral Peak of Front Vowels: A Perceptual Study of the Role of the Second and Third Formants. *Language and Speech*, *10*(3), 181–193. https://doi.org/10.1177/002383096701000304

Gordon, M. J. (2013). Investigating chain shifts and mergers. *The Handbook of Language Variation and Change*, 203–219.

Gross, J., & Forsberg, J. (2020). Weak Lips? A Possible Merger of /i:/ and /y:/ in Gothenburg. *Phonetica*, *77*(4), 268–288. https://doi.org/10.1159/000499107

Gussenhoven, C., & Zhou, W. (2013). *Revisiting pitch slope and height effects on perceived duration*. 1365–1369. https://doi.org/10.21437/Interspeech.2013-360

Hadding, K., Hirose, H., & Harris, K. S. (1976). Facial muscle activity in the production of swedish vowels: An electromyographic study. *Journal of Phonetics*, *4*(3), 233–245. https://doi.org/https://doi.org/10.1016/S0095-4470(19)31246-X

Hadding-Koch, K., & Abramson, A. S. (1964). Duration Versus Spectrum in Swedish Vowels: Some Perceptual Experiments2. *Studia Linguistica*, *18*(2), 94–107. https://doi.org/10.1111/j.1467-9582.1964.tb00451.x

Hammarström, G., & Norman, L. (1957). Om den frikativa slutfasen vid de svenska långa vokalerna i, y, u, w.[On the final fricative phase in the Swedish long vowels i, y, u, w.]. *Nordisk Tidsskrift for Tale Og Stemme*, *17*(3).

Hay, J. B., Pierrehumbert, J. B., Walker, A. J., & LaShell, P. (2015). Tracking word frequency effects through 130 years of sound change. *Cognition*, *139*, 83–91.

Henton, C. G. (2005). Creak as a sociophonetic marker. *The Journal of the Acoustical Society of America*, *80*(S1), S50–S50. https://doi.org/10.1121/1.2023837

Hillenbrand, J. M., Clark, M. J., & Nearey, T. M. (2001). Effects of consonant environment on vowel formant patterns. *The Journal of the Acoustical Society of America*, *109*(2), 748–763. https://doi.org/10.1121/1.1337959

Hillenbrand, J. M., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of american english vowels. *Journal of the Acoustical Society of America*, *97*(5), 3099–3111.

Hillenbrand, J. M., & Nearey, T. M. (1999). Identification of resynthesized /hVd/ utterances: Effects of formant contour. *The Journal of the Acoustical Society of America*, *105*(6), 3509–3523. https://doi.org/10.1121/1.424676

Jacewicz, E., & Fox, R. A. (2018). Regional variation in fundamental frequency of American English vowels. *Phonetica*, *75*(4), 273–309. https://doi.org/10.1159/000484610

Jenkins, J. J., Strange, W., & Miranda, S. (1994). Vowel identification in mixed-speaker silent-center syllables. *The Journal of the Acoustical Society of America*, *95*(2), 1030–1043. https://doi.org/10.1121/1.410014

Johnson, K. (2005). Speaker normalization in speech perception. In D. B. Pisoni & R. E. Remez (Eds.), *The Handbook of Speech Perception* (pp. 363–389). John Wiley & Sons, Inc.

Johnson, K., & Sjerps, M. J. (2021). Speaker normalization in speech perception. In J. S. Pardo, L. C. Nygaard, R. E. Remez, & D. B. Pisoni (Eds.), *The handbook of speech perception* (pp. 145–176). John Wiley & Sons, Inc. https://doi.org/10.1002/9781119184096.ch6

Joos, M. (1948). Acoustic Phonetics. *Language*, *24*(2), 5–136. https://doi.org/10.2307/522229

Kotsinas, U.-B. (1995). *Dialekt [Dialect]* (pp. 267–269). Höganäs: Nationalencyklopedin.

Kuronen, M. (2000). *Vokaluttalets akustik i sverigesvenska, finlandssvenska och finska [The acoustics of vowel pronunciation in Sweden Swedish, Finland Swedish and Finnish]*. Jyväskylä: University of Jyväskylä.

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

Labov, W., Ash, S., & Boberg, C. (2005). *The atlas of north american english: Phonetics, phonology, and sound change*. De Gruyter Mouton.

Ladefoged, P., & Broadbent, D. E. (1957). Information conveyed by vowels. *Journal of the Acoustical Society of America*, *29*, 98–104.

Leinonen, K., Pitkänen, Antti J., & Vihanta, V. V. (1981). Rikssvenskt och finlandssvenskt ljudsystem ur perceptionssynpunkt [Perceptual perspectives on the sound systems of Standard Swedish and Finland Swedish]. *X Fonetiikan päivät, TaYSYLJ 7*, 163–218. Tampere, Finland: Department of Finnish language and general linguistics University of Tampere.

Leinonen, T. (2010). An Acoustic Analysis of Vowel Pronunciation in Swedish Dialects. *Groningen Dissertations in Linguistics*, *83*.

Liljencrants, J., & Lindblom, B. (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language*, 839–862.

Lindblom, B. (1963). *On vowel reduction*. Uppsala: Uppsala University.

Łubowicz, A. (2011). Chain shifts. *The Blackwell Companion to Phonology*, 1–19.

Maclagan, M., & Hay, J. (2007). Getting fed up with our feet: Contrast maintenance and the new zealand english "short" front vowel shift. *Language Variation and Change*, *19*(1), 1–25.

Martinet, A. (1952). Function, structure, and sound change. *WORD*, *8*(1), 1–32. https://doi.org/10.1080/00437956.1952.11659416

McAllister, R., Lubker, J., & Carlson, J. (1974). An EMG study of some characteristics of the Swedish rounded vowels. *Journal of Phonetics*, *2*(4), 267–278. https://doi.org/10.1016/S0095-4470(19)31297-5

Mennen, I., Schaeffler, F., & Docherty, G. (2012). Cross-language differences in fundamental frequency range: A comparison of English and German. *The Journal of the Acoustical Society of America*, *131*(3), 2249–2260. https://doi.org/10.1121/1.3681950

Nearey, T. M. (1978). *Phonetic Feature Systems for Vowels*. Indiana.

Nearey, T. M., & Assmann, P. F. (1986). Modeling the role of inherent spectral change in vowel identification. *The Journal of the Acoustical Society of America*, *80*(5), 1297–1308. https://doi.org/10.1121/1.394433

Pelzer, J. A., & Boersma, P. (2019). Diphthongization in three regional varieties of Swedish. *Proceedings of the 19th International Congress of Phonetic Sciences*, 1144–1148. Canberra, Australia: Australian Speech Science and Technology Association.

Persson, A., Barreda, S., & Jaeger, T. F. (2024). *Comparing accounts of formant normalization against US english listeners' vowel perception*. manuscript; Stockholm University.

Persson, A., & Jaeger, T. F. (2023). Evaluating normalization accounts against the dense vowel space of central swedish. *Frontiers in Psychology*, *14*, 01–21. https://doi.org/10.3389/fpsyg.2023.1165742

Peterson, G. E. (1961). Parameters of Vowel Quality. *Journal of Speech and Hearing Research*, *4*(1), 10–29. https://doi.org/10.1044/jshr.0401.10

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, *24*(2), 175–184.

R Core Team. (2023). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from

https://www.R-project.org/

Renwick, M. E. L., & Stanley, J. A. (2020). Modeling dynamic trajectories of front vowels in the american south. *The Journal of the Acoustical Society of America*, *147*(1), 579–595. https://doi.org/10.1121/10.0000549

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

RStudio Team. (2020). *RStudio: Integrated development environment for r*. Boston, MA: RStudio, PBC. Retrieved from http://www.rstudio.com/

Schaeffler, F. (2005). *Phonological quantity in Swedish dialects: Typological aspects, phonetic variation and diachronic change*. Umeå: Umeå University, Dep. of philosophy and linguistics.

Schertz, J. (2013). Exaggeration of featural contrasts in clarifications of misheard speech in english. *Journal of Phonetics*, *41*(3-4), 249–263.

Schertz, J., & Clare, E. J. (2020). Phonetic cue weighting in perception and production. *WIREs Cognitive Science*, *11*(2), e1521. https://doi.org/https://doi.org/10.1002/wcs.1521

Schötz, S., Frid, J., & Löfqvist, A. (2011). Exotic vowels in swedish – an articulographic and acoustic pilot study of /i:/. *Proc. 17th ICPhS*. Hong Kong.

Schwartz, J.-L., Boë, L.-J., Vallée, N., & Abry, C. (1997). Major trends in vowel system inventories. *Journal of Phonetics*, *25*(3), 233–253.

Seyfarth, S., Buz, E., & Jaeger, T. F. (2016). Dynamic hyperarticulation of coda voicing contrasts. *The Journal of the Acoustical Society of America*, *139*(2), EL31–EL37.

Sóskuthy, M. (2021). Evaluating generalised additive mixed modelling strategies for dynamic speech analysis. *Journal of Phonetics*, *84*. https://doi.org/10.1016/j.wocn.2020.101017

Sóskuthy, M., Foulkes, P., Hughes, V., & Haddican, B. (2018). Changing words and sounds: The roles of different cognitive units in sound change. *Topics in*

Cognitive Science, *10*(4), 787–802. https://doi.org/https://doi.org/10.1111/tops.12346

Ståhle, C. I. (1965). 'Mötet uppnas på sundag'[The meeting uppnas on sundag]. *Språkvård*, *3*(3-8), 1–15.

Stålhammar, U., Karlsson, I., & Fant, G. (1973). Contextual effects on vowel nuclei. *STL-QPSR*, *14*(4), 001–018.

Stevens, K. N., & House, A. S. (1963). Perturbation of vowel articulations by consonantal context: An acoustical study. *Journal of Speech and Hearing Research*, *6*(2), 111–128. https://doi.org/10.1044/jshr.0602.111

Stevens, M., Harrington, J., & Schiel, F. (2019). Associating the origin and spread of sound change using agent-based modelling applied to /s/-retraction in English. *Glossa: A Journal of General Linguistics*, *4*(1), 01–30. https://doi.org/10.5334/gjgl.620

Stilp, C. (2020). Acoustic context effects in speech perception. *WIREs Cognitive Science*, *11*(1), 1–18. https://doi.org/10.1002/wcs.1517

Strange, W. (1989). Evolving theories of vowel perception. *The Journal of the Acoustical Society of America*, *85*(5), 2081–2087. https://doi.org/10.1121/1.397860

Strangert, E. (2001). Quantity in ten swedish dialects in northern sweden and Österbotten in finland. *Working Papers/Lund University, Department of Linguistics and Phonetics*, *49*, 144–147.

Syrdal, A. K. (1985). Aspects of a model of the auditory representation of american english vowels. *Speech Communication*, *4*(1-3), 121–135. https://doi.org/10.1016/0167-6393(85)90040-8

Watson, C. I., & Harrington, J. (1999). Acoustic evidence for dynamic formant trajectories in Australian English vowels. *The Journal of the Acoustical Society of America*, *106*(1), 458–468. https://doi.org/10.1121/1.427069

Wedel, A., Nelson, N., & Sharp, R. (2018). The phonetic specificity of contrastive hyperarticulation in natural speech. *Journal of Memory and Language*, *100*, 61–88.

Weirich, M., Simpson, A. P., Öjbro, J., & Ericsdotter Nordgren, C. (2019). The phonetics of gender in Swedish and German. *Fonetik 2019, Stockholm, Sweden, 10-12 June, 2019*, 49–53.

Wenner, L. (2010). *När lögnare blir lugnare. En sociofonetisk studie av sammanfallet mellan kort ö och kort u i uppländskan [When lögnare become lugnare. A sociophonetic study of the merger between short ö and short u in Uppland Swedish]*. Uppsala: Uppsala Universitet.

Wieling, M. (2018). Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of english. *Journal of Phonetics*, *70*, 86–116.

Wright, R., Local, J., Ogden, R., & Temple, R. (2004). Factors of lexical competition in vowel articulation. *Papers in Laboratory Phonology VI*, 75–87.

Xie, X., & Jaeger, T. F. (2020). Comparing non-native and native speech: Are L2 productions more variable? *The Journal of the Acoustical Society of America*, *147*(5), 3322–3347. https://doi.org/10.1121/10.0001141

Young, N. J., & McGarrah, M. (2021). Forced alignment for Nordic languages: Rapidly constructing a high-quality prototype. *Nordic Journal of Linguistics*, 1–27. https://doi.org/10.1017/S033258652100024X