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 on static and dynamic acoustic analyses of the entire vowel space using a recently released database of *h-VOWEL-d* words (SwehVd). The results highlight the importance of static and dynamic spectral and temporal cues for Swedish vowel category distinction. The first two formants and vowel duration are the primary acoustic cues to vowel identity, however, the third formant contributes to increased category separability for neighboring contrasts presumed to differ in lip-rounding. In addition, even though all long-short vowel pairs differed systematically in duration, they also display considerable spectral differences, suggesting that quantity distinctions are not separate from quality distinctions in Swedish. The dynamic analysis further suggests formant movements in both long and short vowels, with [e:] and [o:] displaying clearer patterns of diphthongization.

Keywords: vowels, category separability, formant dynamics

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 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), formant movements for some front contrasts (Kuronen, 2000; Pelzer & Boersma, 2019), as well as 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). Single-point estimates extracted from the steady-state of the vowel where the formant pattern is presumed to be static, continues to be widely used to represent vowels in an F1-F2 plane (for a review, see Kent & Vorperian, 2018). However, static estimates cannot capture how formants move as the signal unfolds; information that has been shown to influence listeners' vowel perception (e.g., Assmann & Katz, 2005; Eklund & Traunmüller, 1997; Hillenbrand & Nearey, 1999; Kuronen, 2000; Nearey & Assmann, 1986).

This paper investigates the acoustic characteristics of modern-day Swedish vowels in analyses that aim to contribute to our understanding of language-specific and language-general patterns of vowel acoustics. The paper presents a comprehensive description of the primary acoustic cues to vowel identity, using a recently released database of *h-VOWEL-d* (short: hVd) words, the SwehVd database (Persson & Jaeger, 2023). 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; Leinonen, 2010; Riad, 2014). Two of the most recent studies are 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 (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), allophonic variation in /ɛ/ (Gross, Boyd, Leinonen, & Walker, 2016), or a single allophone, 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

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

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.

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
Gross, Boyd, Leinonen & Walker, 2016	3-15 repetitions of /ɛː/ before retroflex consonants or in other word contexts extracted from running speech (alongside corner vowel productions of [iː], [ɑː], [uː])	57 female and male talkers of 16-17 years of age (28 from Stockholm; 13 with foreign-born mother)	Euclidean distance; Wilcoxon rank sum test; t-test	Sociolinguistic variation in allophones to /ɛː/ in Gothenburg and Stockholm Swedish
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 consonantal 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 paper provides an up-to-date acoustic description of the entire vowel space of 21 categories, using the SwehVd database (Persson & Jaeger, 2023). The hVd context continues to be widely used in vowel production and perception studies on languages where the glottal /h/ in onset minimizes supraglottal articulations and thereby reduces the risk of coarticulatory effects from the surrounding phonetic context, as in e.g., English and Swedish (as confirmed in e.g., Chesworth, Coté, Shaw, Williams, & Hodge, 2003; Robb & Chen, 2009). The use of an hVd database hence increases comparability to studies on other languages (e.g., Hillenbrand, Getty, Clark, & Wheeler, 1995; Peterson & Barney, 1952). The main spectral (F1-F2-F3) 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). 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 and is therefore reported in the static analysis for a comprehensive picture of the acoustics (e.g., Henton, 2005; Jacewicz & Fox, 2018; Johnson, 2005; Leung, Jongman, Wang, & Sereno, 2016; Mennen, Schaeffler, & Docherty, 2012; Weirich, Simpson, Öjbro, & Ericsdotter Nordgren, 2019).

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; Persson & Jaeger, 2024), 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.

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 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] - [ɵ], 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; Leinonen, 2010).[2]

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

Previous studies have characterized the diphthongal glide in the later part of the long 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 buzzing sound is presumably generated by the co-articulation of the vowel and a voiced fricative sound similar to [z] (Elert, 1980; Engstrand, Björsten, Lindblom, Bruce, & Eriksson, 2000). 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 et al., 2000; Gross & Forsberg, 2020; Riad, 2014). It has been claimed to carry strong socio-indexical meaning across locations; indexing place in rural areas, and class and gender in urban areas (e.g., Bruce, 2010; Gross, 2018; Kotsinas, 1994; Nilsson, Wenner, Leinonen, & Thorselius, 2021). 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.

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

2 Methods

2.1 Materials

The materials used 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, the final release is presented, including 24 additional male talkers. 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 10 recordings of each hVd word from each of the 48 talkers (N = 24 female), for a total N of

[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]. The mix of 4 real Swedish words – *hed*, *härd*, *hörd*, *hud* (English translations: *heath*, *hearth*, *heard* and *skin*, respectively), and 18 phonotactically legal pseudowords in the word list, might have affected talkers' pronunciations. For instance, there are studies indicating frequency and neighborhood density effects on vowel productions, with low-frequency words and words with high neighborhood density being produced more distinctly (e.g., Munson & Solomon, 2004; Wright, Local, Ogden, & Temple, 2004). Similar effects of hyperarticulation interacting with neighborhood density have been found for pseudowords as well, however, with substantial cross-talker variability (e.g., Scarborough, 2012).

tokens = 9979. 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 = 29; SD = 6.79). 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 41 L1 talkers (N=20 female talkers), with on average 359 (SD = 21.50) tokens per vowel (range = 304 to 383), for a total of 7529 observations.

2.2 Acoustic analyses

2.2.1 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, 1992-2022), using the Burg algorithm to extract

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

estimates of the first three formants (F1-F3) at five time-points of the vowel (20, 35, 50, 65, 80% into the vowel). The five points were selected to capture formant trajectories and potential diphthongization in the later time-points for the dynamic analysis (following, e.g., Holbrook & Fairbanks, 1962; Jacewicz, Fox, & Salmons, 2011; Kuronen, 2000; Lehiste & Peterson, 1961; Yang, 2019), as well as provide stable measures of the steady-state of the vowel for the static analysis. 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. Measures of vowel duration and the mean F0 were extracted across the entire vowel segment. The Praat script that extracts these cues is shared as part of the SwehVd OSF repository, allowing researchers to choose additional or alternative time points at which to extract formants and F0.

To correct for measurement errors in the automatic extraction of cues, 5 separate univariate distributions of the five extracted cues (F0, F1, F2, F3 and duration) was estimated for each distinct combination of talker and vowel. Points that fell outside of the 2.50th to the 97.50th quantile of the distributions for each vowel were identified, examined for measurement errors, and subsequently corrected. This followed the approach employed for the SwehVd corpus (Persson & Jaeger, 2023), and strikes a middle-ground between the ideal (manual correction of all tokens) and feasibility.

2.2.2 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 S1.5. Vowel normalization is used in studies on vowel production and perception to account for acoustically irrelevant cross-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).

2.2.3 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] 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.

In order to evaluate the hypothesized importance of lip-rounding (F3) for neighboring unrounded and rounded categories, a category separability index was employed. Category separability index and similar measures of reduction in variance or distance between means continue to be frequently used in phonetic research (see e.g., Nycz & Hall-Lew, 2013 for a review; Fabricius, Watt, & Johnson, 2009; Flynn & Foulkes, 2011; Labov, 2010). While the

[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 (S1.3) presents evaluations of the effect of different measurement points.

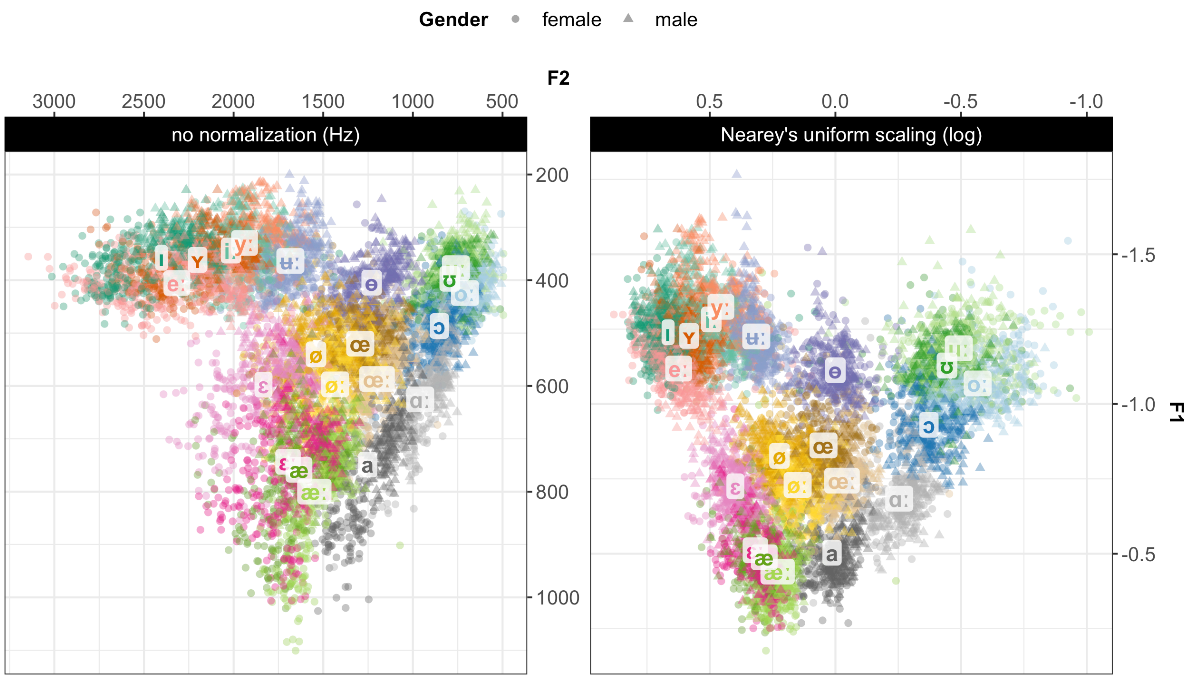

Figure 1. The SwehVd vowel data in unnormalized Hertz (*left*) and Nearey's uniform scaling space (*right*), along the first two formants, F1 and F2. Points show recordings of each of the 21 Central Swedish vowels by 44 (24 female) L1 talkers in the database, averaged across the three middle time-points (at 35, 50, 65% into the vowel). Vowel labels are placed at the vowel mean across talkers. Long vowels are boldfaced. Vowel recordings on which the talker produced a different vowel than the intended are excluded (1.21% of all recordings).

separability index is a simple and straightforward measure of the relative separability of vowel categories, it nevertheless comes with certain limitations to which I return in the General discussion (4). Following work by Wedel, Nelson, and Sharp (2018) and 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).

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

For instance, for the [iː] - [yː] 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 same was done in the opposite direction for the same contrast, thus the calculating the separability of [iː] tokens from the [yː] center, following Xie and Jaeger (2020). The separability index reports the average separability across the two categories in each contrast. 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 [iː] - [yː], [eː] - [yː], [ɪ] - [ʏ] for comparing unrounded vs. outrounded vowels, and [yː] - [ʉː], [oː] - [uː], [ɔ] - [ʊ] for outrounded vs. inrounded vowels.[6]

To quantify the effect of including F3 on category separability, separate linear mixed-effects models (LMM) were fit for each contrast, predicting separability from cue combination (F1-F2-F3 vs. F1-F2) while including by talker random intercepts.[7] The model was formulated as follows: $separability \sim cuecombination + (1|Talker)$. Cue combination was treatment-coded with F1-F2 as reference category, thus comparing F1-F2-F3 against the F1-F2 baseline.

The same process was applied to investigate to what extent long-short vowel pairs differ in spectral and temporal cues, by assessing what combination of cues would 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 three different cue combinations: F1-F2, F1-F2-F3, or F1-F2-duration. The models were the same as the previous sets, with F1-F2 as reference category, comparing each cue combination against

[6] I note that this way of calculating category separability assumes talker-specific category representations. The SI S1.7.3 reports separability indices that instead assume talker-independent representations.

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

the F1-F2 baseline.

The results of the static analysis are presented in Section 3.1.

2.2.4 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 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 6 subsets of neighboring contrasts, hypothesized to differ primarily in formant dynamics: [iː] - [yː], [iː] - [ʉː], [iː] - [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 employed ordered factor difference smooths as follows: $formant \sim category + Gender + s(timepoint) + s(timepoint, by = category, k = 5) + s(Talker, bs = "re") + s(Talker, category, bs = "re")$. The ordered factor predictor (= category) was treatment coded with [iː], [oː], [ɛː], [øː], [ɪ], [ɔ], [ɛ], and [ø] as reference categories in respective set.

The second group consisted of 11 sets of GAMMs fit to each of the long-short vowel pairs, aiming for an evaluation of differences between categories within pairs. In each set, vowel was treatment coded with the long vowel in each pair as reference vowel.

All GAMMs were fit separately for each of the three 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 3.2.

3 Results

This study 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 results section is divided into two main sections, following the static and dynamic analyses.

3.1 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.

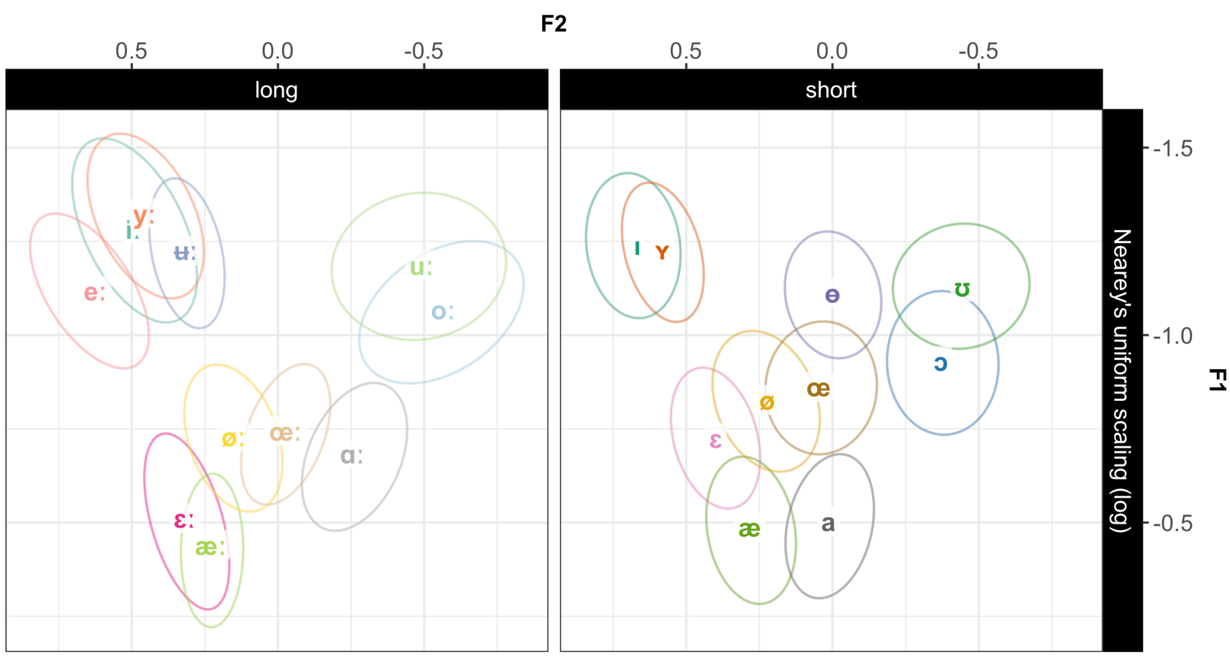

Figure 2. The SwehVd vowel data separated by quantity. Ellipses show bivariate Gaussian 95% confidence interval of vowel means. Vowel labels indicate vowel means across female and male talkers.

Figure 2, left panel, visualizes the long vowels along the two primary cues to vowel identity, F1-F2.[8] Four vowels cluster in the high front part of the space. The mid-high [eː] 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 each other and with the neighboring [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).[9] 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 and more peripheral than their long counterparts, which does not replicate previous studies (for Central Swedish, see e.g., Fant, 1971; Kuronen, 2000; for short-long contrasts in other languages, see e.g., Clopper, Pisoni, & De Jong, 2005; Hillenbrand et al., 1995). There are further no indications of neutralization of the short /ø/ and /ʉ/ as [ɵ] among these talkers (c.f., Ståhle, 1965; Wenner, 2010).

3.1.1 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 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.

[8] The SI presents the mean cue values for the male and female talkers, Tables S1 and S2. As expected, the male talkers have lower formant values and lower F0s than the female talkers (average F0 across long and short categories for female talkers = 204, for male talkers = 119).

[9] 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.

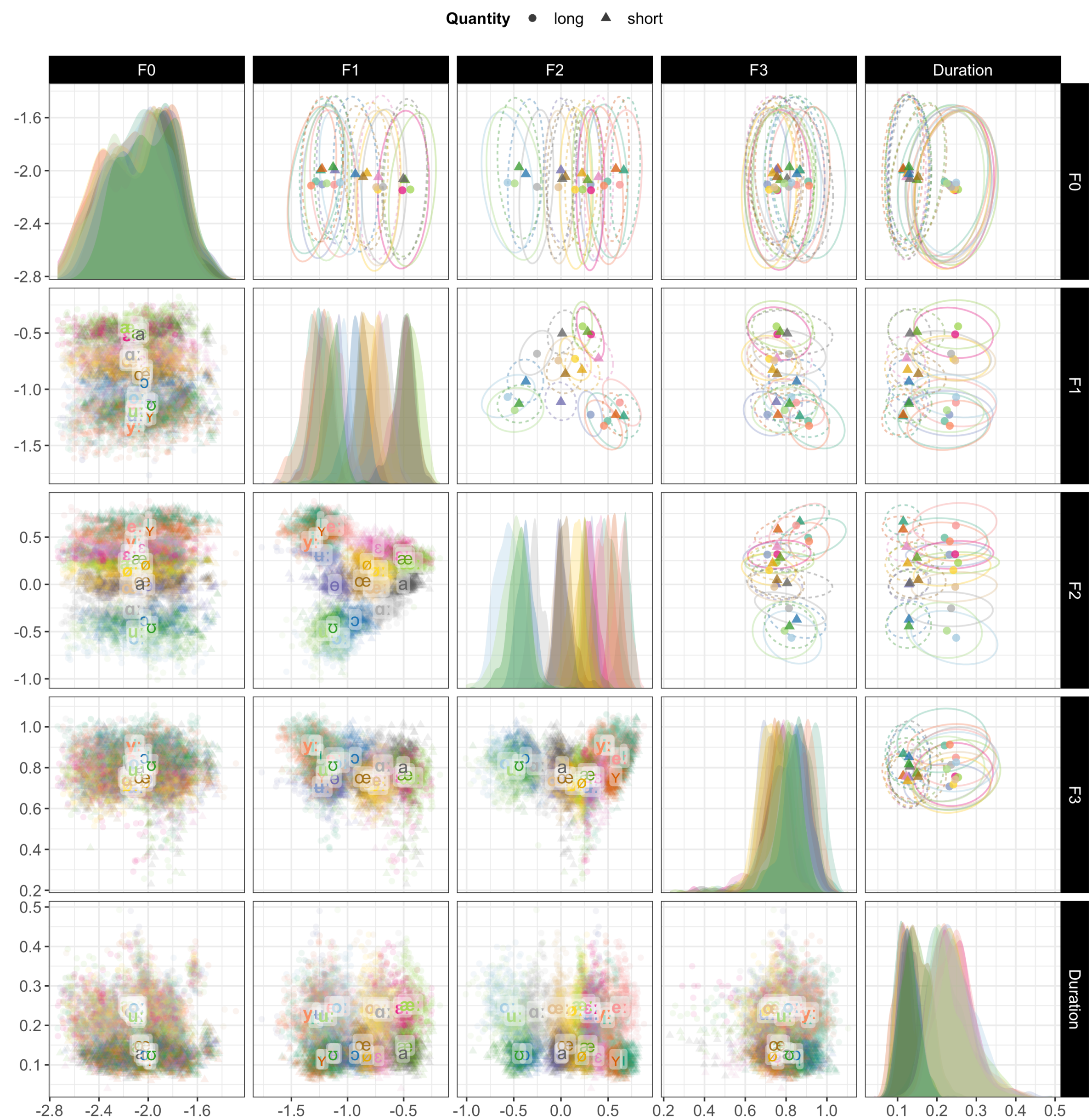

Figure 3. The SwehVd vowel data shown for all pairwise combinations of five cues: F0, F1, F2, F3 and duration. Panels on the diagonal show marginal cue densities of all five cues. The off-diagonal panels show vowel means across talkers, represented by points and with bivariate Gaussian 95% probability 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.

As is to be expected, vowels differing in quality are most separated in the F1 and F2 panels. Inspecting the off-diagonals in Figure 3, 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 individual differences in the amount of overlap exist. Most talkers produce these two vowels very close in F1-F2 space, and only 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.

Three general observations can be made from Figure 4. First, category separability is overall lower for some contrasts when only F1 and F2 were considered, e.g., the [iː] - [yː], and [ɪ] - [ʏ] contrasts, indicating their overlap in F1-F2 space (Figure 4:**column A**). Second, including F3 overall increases category separability, more so for some contrasts than others. The proportional increase in category separability relative to F1-F2 baseline (Figure 4:**column B**) is largest for the [iː] - [yː], [ɪ] - [ʏ], and [yː] - [ʉː] contrasts. How much separability increases by adding F3 varies quite substantially between talkers for the [iː] - [yː] and [yː] - [ʉː] contrasts, as indicated by the large confidence intervals. Indeed, when assuming talker-independent representations, the relative increase in separability is less pronounced for these contrasts (see additional analyses in SI S1.7.3). Third, the [yː] - [ʉː] contrast seems to benefit most from the inclusion of F3, resulting in an overall larger increase in separability for the outrounded vs. inrounded contrasts over the unrounded vs. outrounded contrasts.

The LMMs fit to the data (presented in Section 2.2.3) indicated that including F3 improved category separability for all contrasts (all $ps < .003$). This suggests that the subtle differences observed by visual inspection for the [eː] - [yː], [oː] - [uː] and [ɔ] - [ʊ] contrasts were nevertheless significant (Summary tables in SI S1.7.1).[10]

3.1.2 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.

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 database, which increases the distance

[10] The alpha level for statistical significance used throughout the paper is $p < .05$.

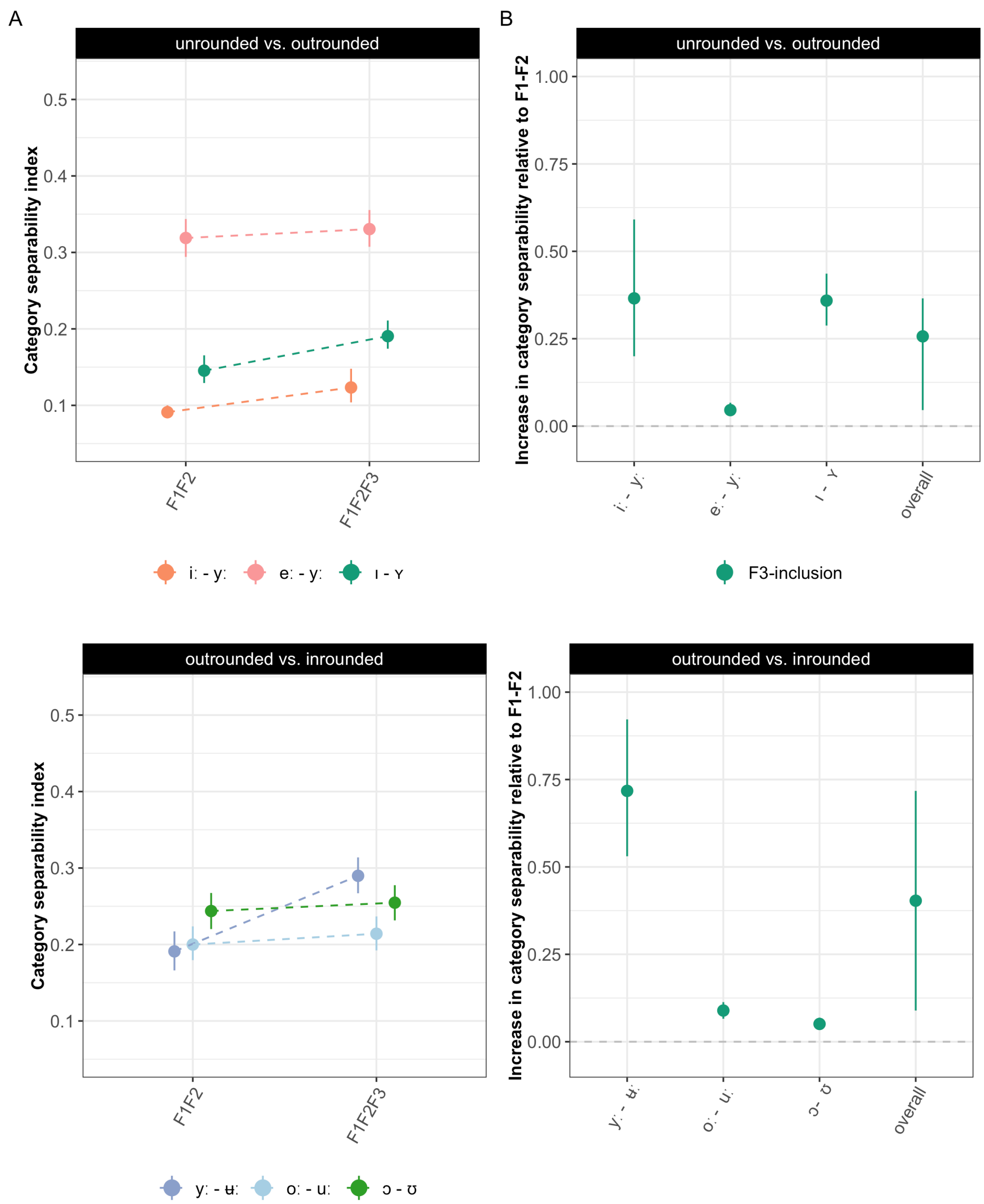

Figure 4. The effect of including F3 in measures of category separability for the distinction between neighboring unrounded vs. outrounded vowels (**top row**), and outrounded vs. inrounded vowels (**bottom row**). **Left panels** plot the category separability for F1-F2 and F1-F2-F3 cue combinations. **Right panels** plot the proportional increase in category separability relative to F1-F2-baseline. Pointranges indicate mean and 95% bootstrapped CIs of the category separability summarized across talkers for each cue combination. Axis ranges are held constant across columns.

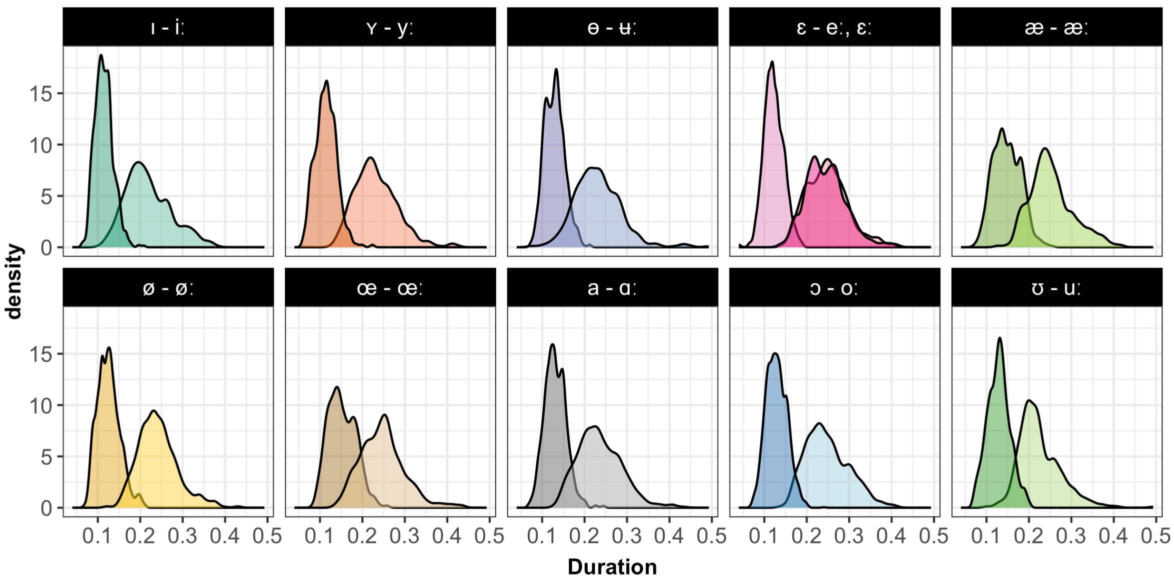

Figure 5. Illustrating the systematic differences in duration between the long and short vowel pairs in SwehVd.

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 three different cue combinations: F1-F2, F1-F2-F3, or F1-F2-Duration.[11]

Figure 6 indicates that category separability is generally higher for some pairs. For instance, the [eː] - [ɛ], [ʉː] - [ɵ], and [ɑː] - [a] pairs display the largest separability when only F1-F2 is considered (Figure 6:**column A**). The inclusion of additional cues unsurprisingly

[11] For an approximation of the relative separability by cue, evaluations of F1-F3, F2-F3, F1-Duration and F2-Duration are included in the SI S1.7.2.

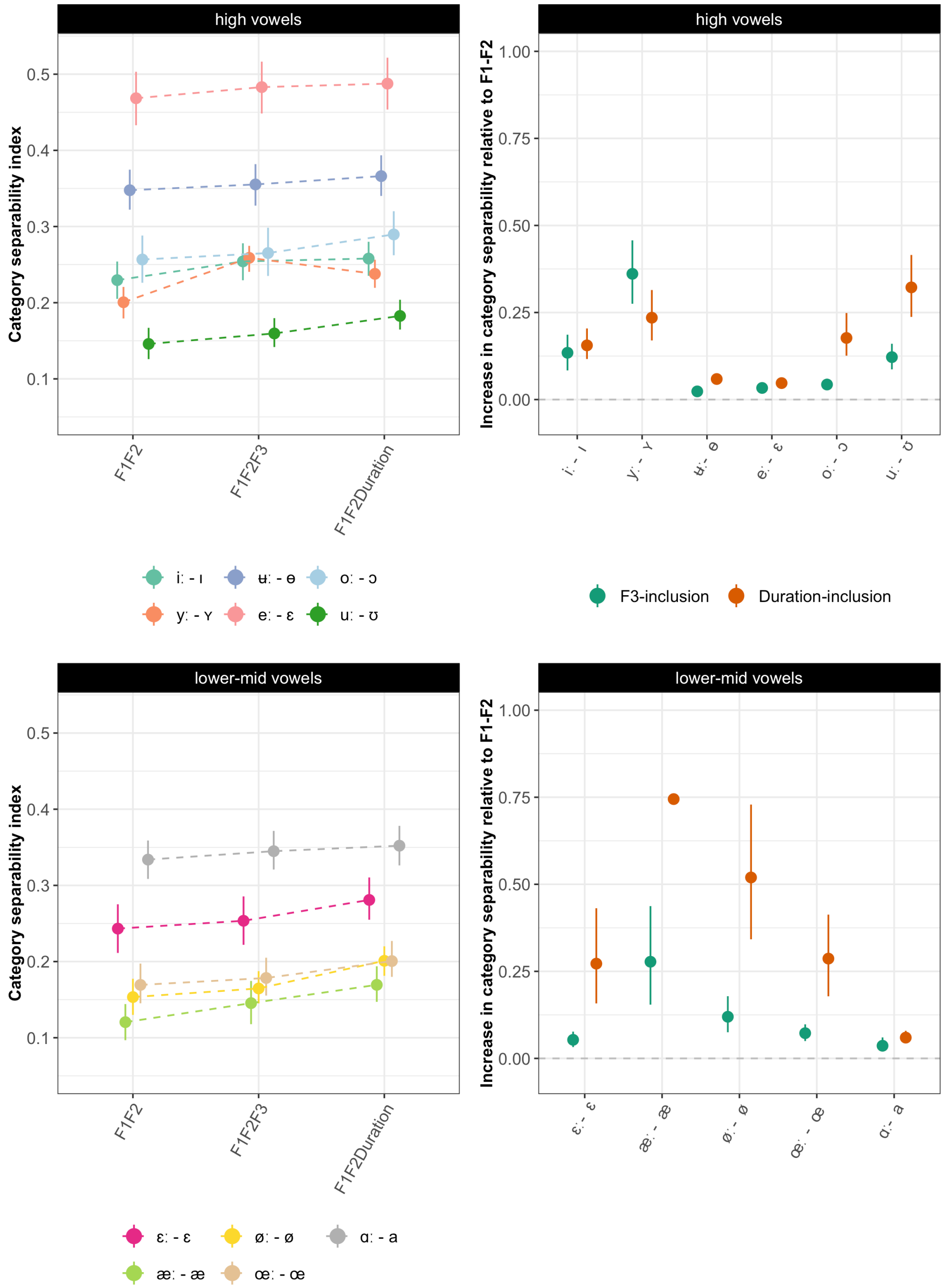

Figure 6. The effect of including F3 and duration in measures of category separability for long-short vowel pair distinctions. For visualization purposes, the pairs are split into high vowels (**top row**), and lower-mid vowels (**bottom row**). **Left panels** plot the category separability for F1-F2, F1-F2-F3 and F1-F2-duration cue combinations. **Right panels** plot the increase in category separability relative to F1-F2-baseline. Pointranges indicate mean and 95% bootstrapped CIs of the category separability summarized across talkers for each cue combination. Axis ranges are held constant across columns.

increases category separability for all pairs, more so for pairs that are less separable in F1-F2 space (e.g., [uː] - [ʊ], [yː] - [ʏ], [æː] - [æ], [øː] - [ø], and [œː] - [œ]). The inclusion of duration overall maximizes the increase in separability relative to baseline (Figure 6:**column B**). Interestingly, this is not the case for [yː] - [ʏ] that achieves the highest separability for the F1-F2-F3 combination. This would overall seem to suggest that the first two formants and duration are the most important cues to long-short vowel pair distinctions, while the inclusion of F3 unsurprisingly does not punish the separability.[12]

These findings were confirmed by the statistical analysis (Summary tables in SI S1.7.1): the LMMs indicated that the F1-F2-Duration cue combination generated the highest separability for all pairs (all $ps > .0001$), with the exception of the [yː] - [ʏ] contrast, for which the F1-F2-F3 combination achieved the highest separability ($\hat{\beta} = .058, SE = .004, p < .0001$). The inclusion of F3 nevertheless increased separability relative to F1-F2 for all pairs (all $ps > .01$), with the [ɛː] - [ɛ] ($\hat{\beta} = .01, SE = .004, p > .009$) and [œː] - [œ] ($\hat{\beta} = .009, SE = .0034, p < .01$) pairs displaying overall smaller but statistically significant differences.

To sum up, the results of the static analysis suggest that F1, F2 and duration 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 some 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

[12] Analyses of additional cue combinations in the SI S1.7.2 (F1-F3, F2-F3, F1-Duration, F2-Duration) suggested that duration contributed more to separability than F3 for the [ɛː] - [ɛ], [æː] - [æ], [øː] - [ø], [œː] - [œ], and [uː] - [ʊ] vowel pairs, while for the [ʉː] - [ɵ], [eː] - [ɛ], [oː] - [ɔ], [ɑː] - [a] contrasts, combining one spectral cue with duration decreased separability relative to baseline, hence highlighting the reliance on both F1-F2 over duration.

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 F1-F2-duration combination generated the highest separability for all pairs but the [yː] - [ʏ], where F3-inclusion maximized separability of the cue combinations considered, highlighting the importance of F3 for this contrast.

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 separability. 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.

3.2 Dynamic spectral analysis

This section 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 dynamics in eight sets of neighboring vowel contrasts, and subsequently on the long and short vowel pairs.

3.2.1 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 very 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].

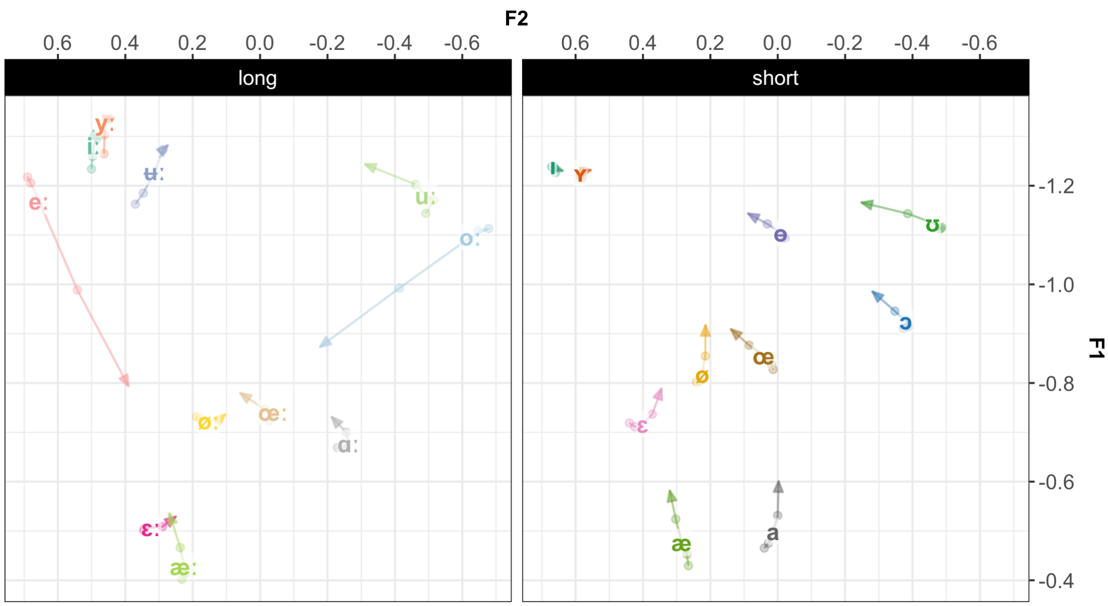

Figure 7. The trajectory of all vowels across the five time-points, along F1-F2. The arrow indicates the direction of the trajectory and ends at the final time-point, at 80% into the vowel. The vowel label is placed at the third time-point, at vowel mid-point (50%). The first (20%), second (35%) and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

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 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; 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.

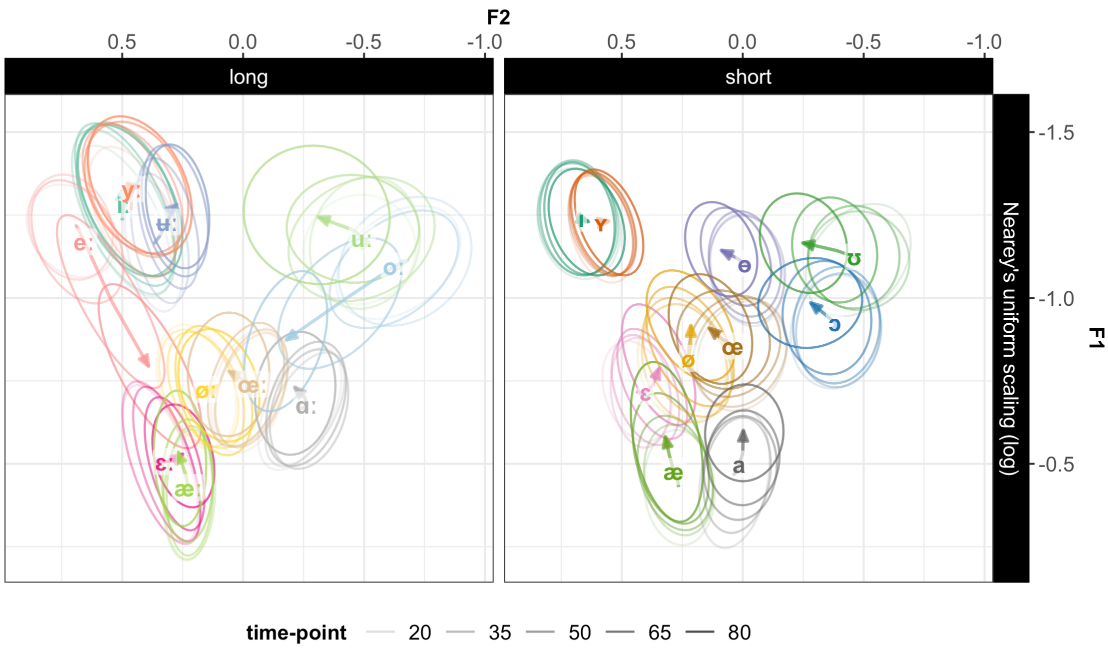

Figure 8. Vowel placement in F1-F2 space at each of the five time-points. Ellipses show bivariate Gaussian 95% confidence interval of vowel means at each of the five time-points. Transparency indicates time-point, more transparent ellipses for earlier times. The vowel label is placed at vowel mid-point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

3.2.2 Models of formant dynamics

3.2.2.1 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 the 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 S1.8.1.

The GAMMs fit to the high front vowel contrasts suggested significant constant and non-linear differences over time between vowels in each contrast for all cues and contrasts (for constant differences, all $ps < .0001$; for non-linear differences, all $ps < .039$), except for [iː] - [yː] predicting F3 ($\hat{\beta} = -.002, SE = .01, p > .84$), and [ɪ] - [ʏ] predicting F1 ($\hat{\beta} = .007, SE = .007, p > .27$) (Figures 9 and 10). This suggests that there are no differences in F3 for [iː] and [yː] or in F1 for [ɪ] and [ʏ] when formant dynamics are considered. While F3 increases separability between [iː] and [yː] under static analysis, the *dynamics* in F3 does not seem to add information about vowel quality for this contrast.

Significant constant as well as non-linear effects of category on all three cues was found in all GAMMs fit to the high back long and short contrasts (for constant differences, all $ps < .0001$; for non-linear differences, all $ps < .0197$). These results suggest that all three cues contribute to distinguishing between these vowels also when formant dynamics are considered (Figure 11).

The GAMMs fit to the lower-mid front long and short vowel contrasts suggested constant and non-linear effects of vowel on F1 and F2 (for constant differences, all $ps < .0001$; for non-linear differences, all $ps < .0002$). For F3, there were no significant constant differences between [ɛː] and [æː] ($\hat{\beta} = -.006, SE = .007, p > .37$), and no constant or non-linear differences between [ɛ] and [æ] ($\hat{\beta} = .0099, SE = .005, p > .073$; $EDF = 1, F = .879, p > .348$). Figure 12 demonstrates how these vowels overlap in F3-dynamics, but are distinguished for most of the segment along F1 and F2.

For the GAMMs fit to the mid center vowels, there was an effect of category on all cue evaluations for both long and short vowels (for constant differences, all $ps < .048$; for non-linear differences, all $ps < .0001$; Figure 13), with the exception of both long and short vowels fit to F1, for which no constant difference was found (for [øː] - [œː], $\hat{\beta} = -.0098, SE = .007, p > .18$; for [ø] - [œ], $\hat{\beta} = -.016, SE = .0097, p > .102$). These contrasts nevertheless displayed non-linear differences over time (Figure 13). These results would seem to suggest that when formant dynamics are considered, F1 does not carry information about vowel quality for the [øː] - [œ] and [ø] - [œː] contrasts. However, the vowels within each pair display different non-linear patterns.

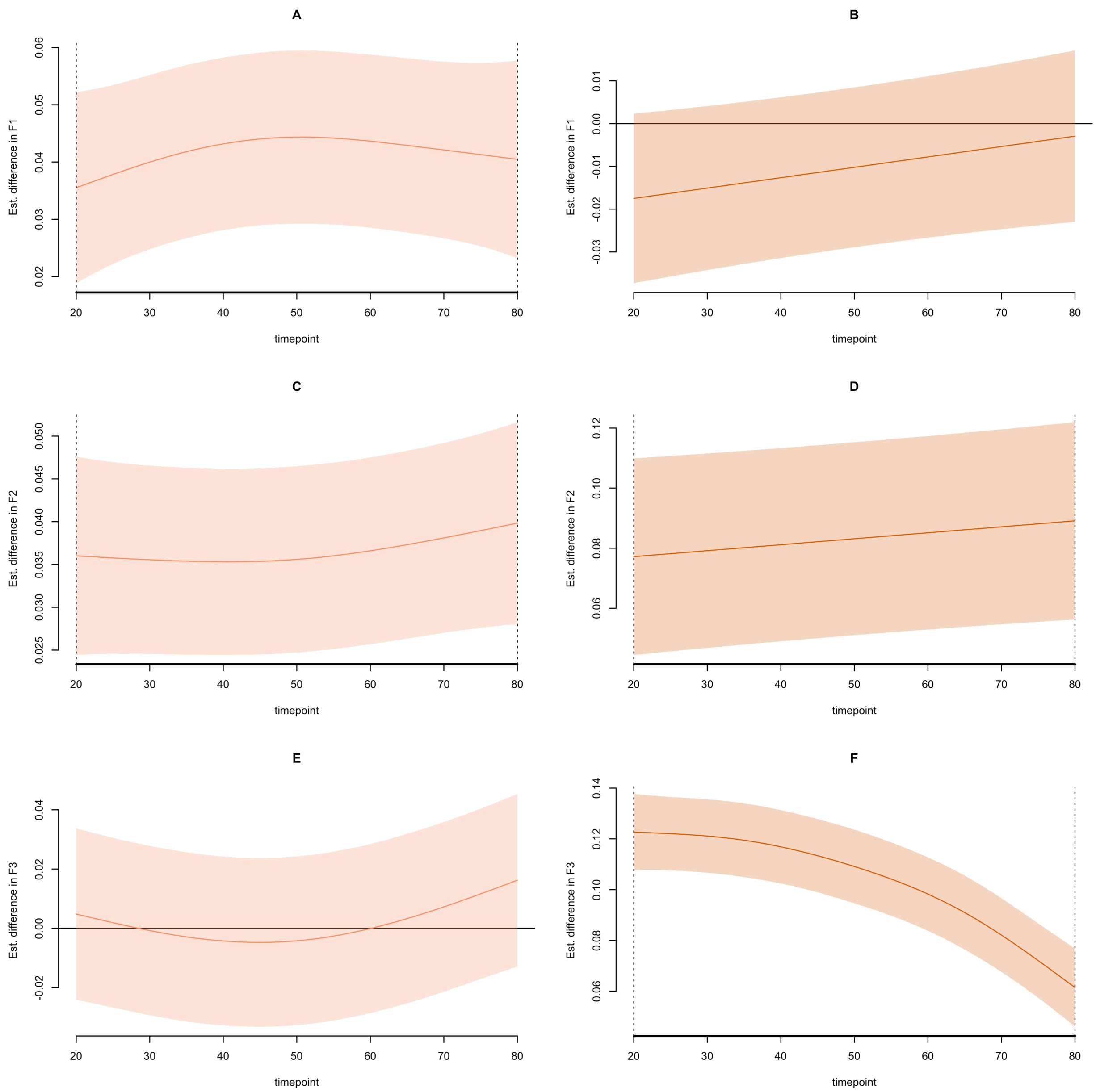

Figure 9. Fitted smooths of GAMM for predicting F1 (**upper row**), F2 (**mid row**), F3 (**bottom row**) and 95% confidence intervals for the [iː] - [yː] contrast (**left**), and the [ɪ] - [ʏ] contrast (**right**). Differences significantly different from 0 are marked by black dotted vertical lines.

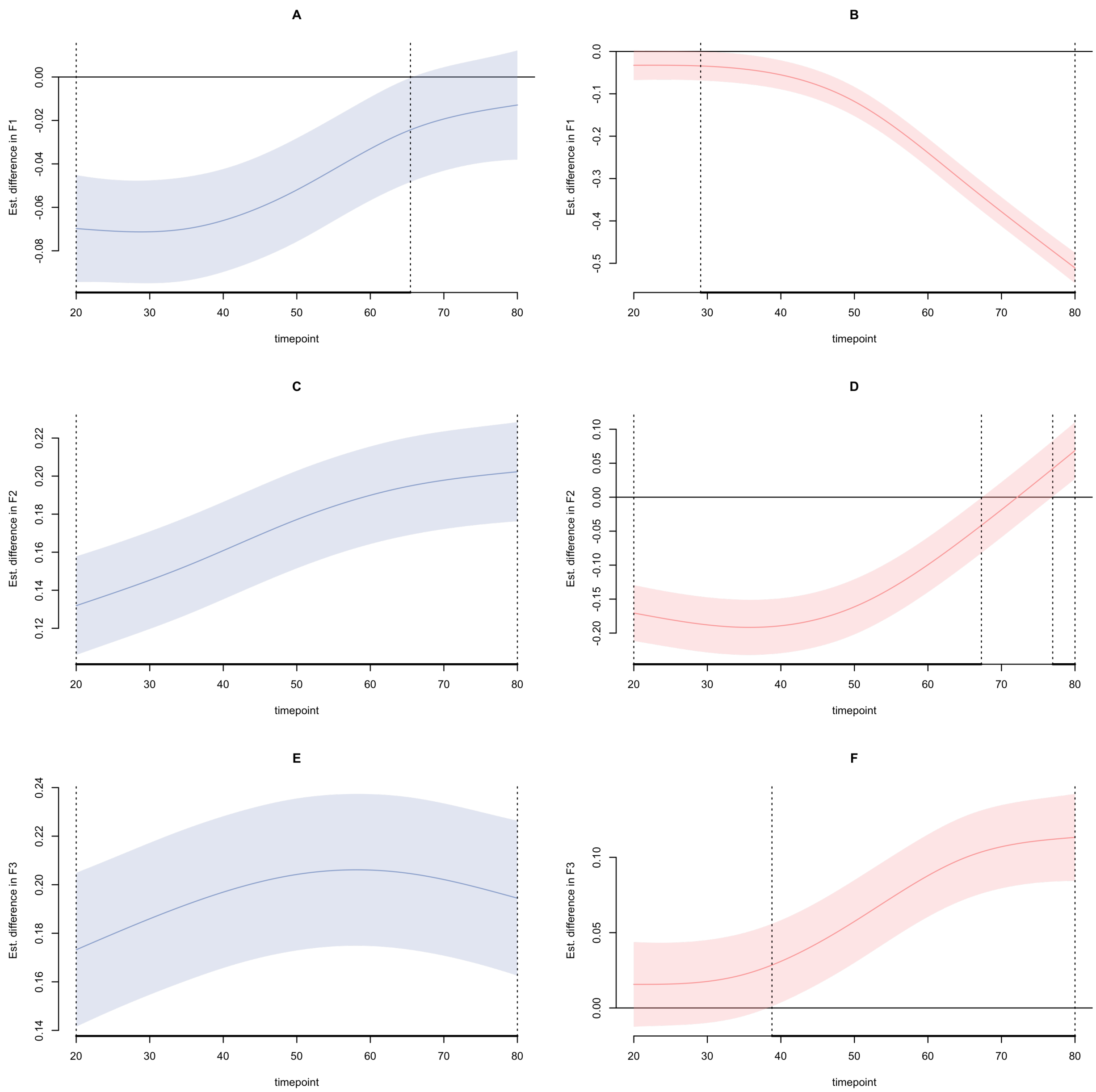

Figure 10. Fitted smooths of GAMM for predicting F1 (**upper row**), F2 (**mid row**), F3 (**bottom row**) and 95% confidence intervals for the [iː] - [ʉː] contrast (**left**), and the [iː] - [eː] contrast (**right**). Differences significantly different from 0 are marked by black dotted vertical lines.

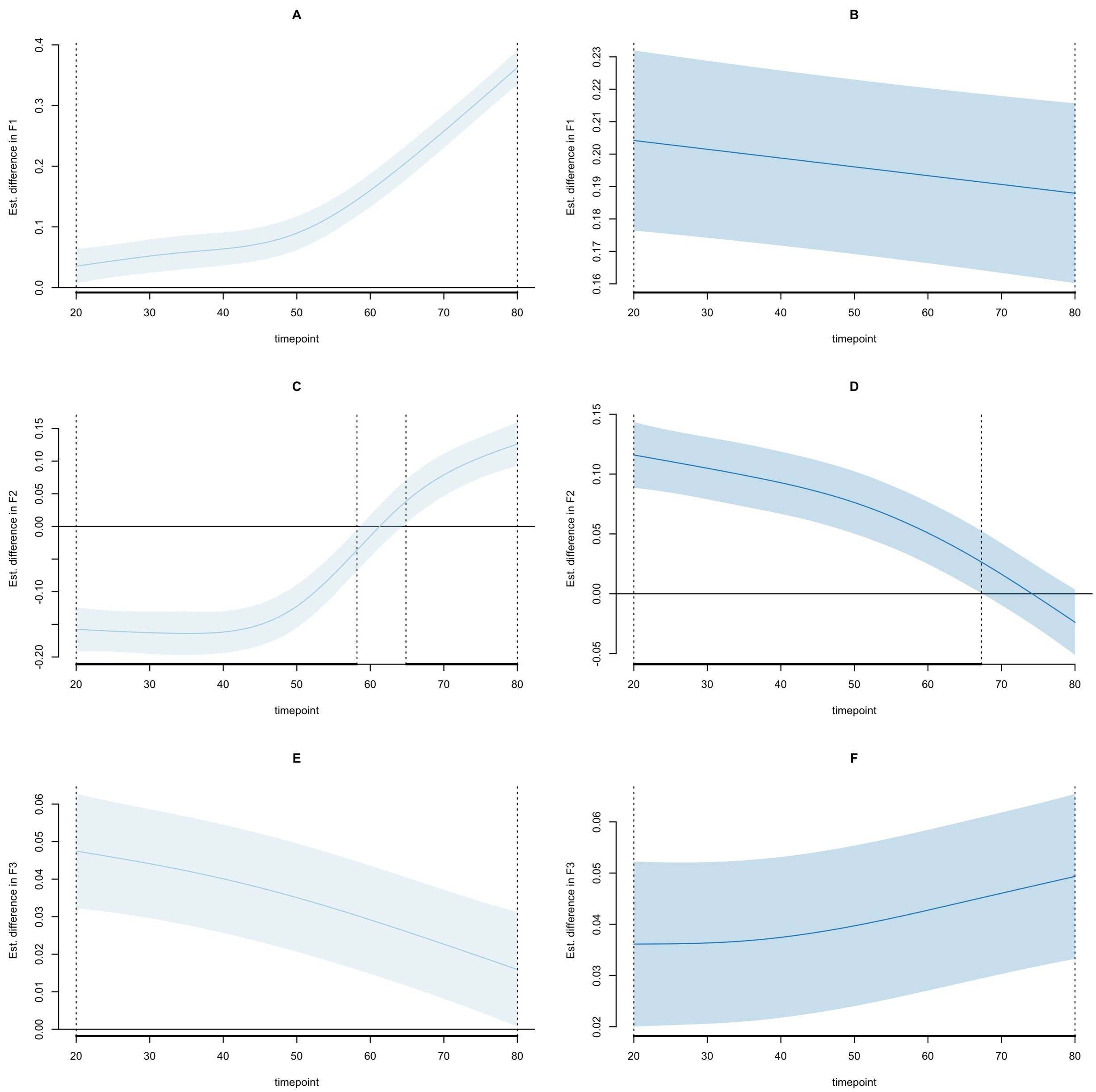

Figure 11. Fitted smooths of GAMM for predicting F1 (**upper row**), F2 (**mid row**), F3 (**bottom row**) and 95% confidence intervals for the [oː] - [uː] contrast (**left**), and the [ɔ] - [ʊ] contrast (**right**). Differences significantly different from 0 are marked by black dotted vertical lines.

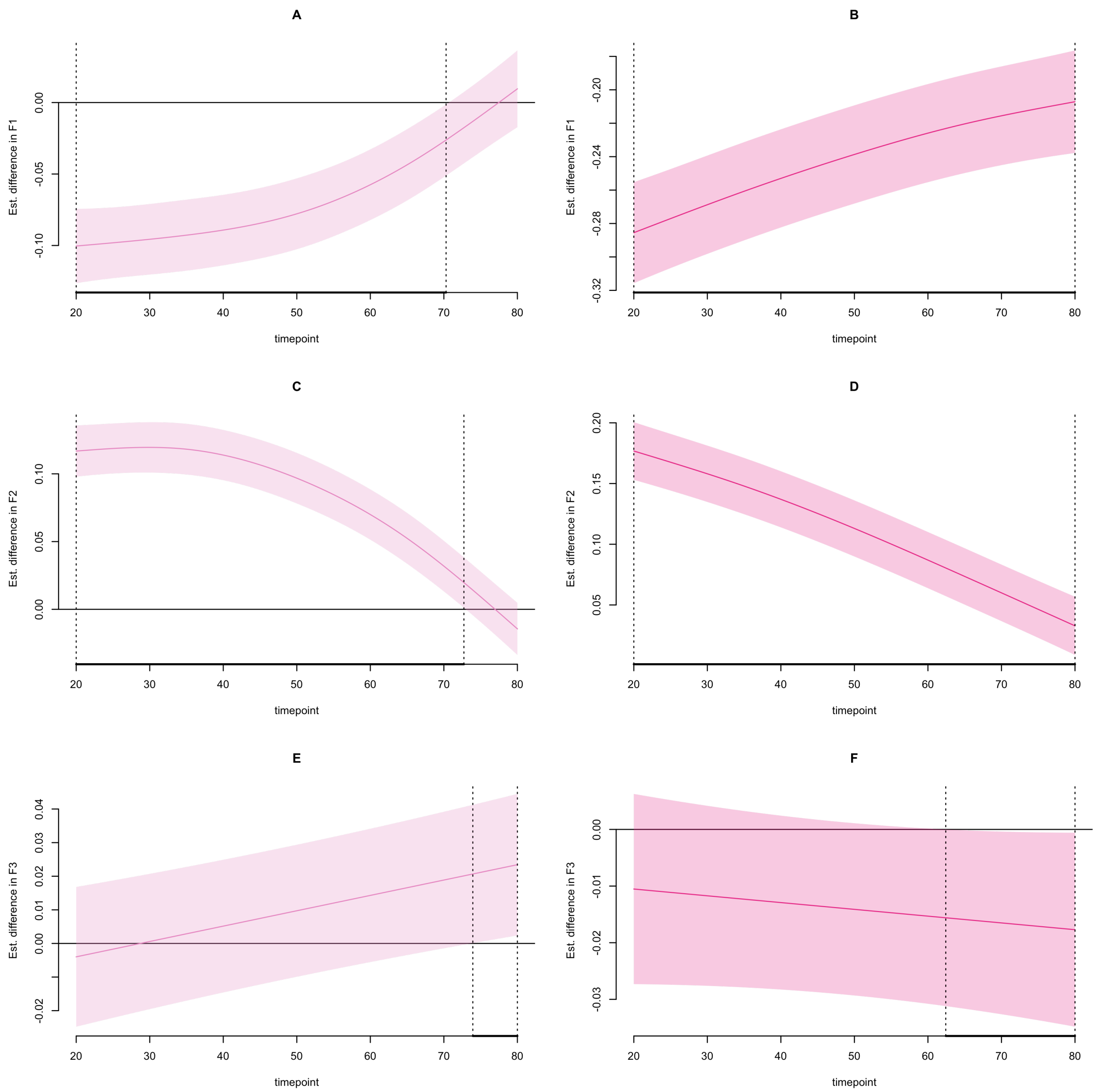

Figure 12. Fitted smooths of GAMM for predicting F1 (**upper row**), F2 (**mid row**), F3 (**bottom row**) and 95% confidence intervals for the [ɛː] - [æː] contrast (**left**), and the [ɛ] - [æ] contrast (**right**). Differences significantly different from 0 are marked by black dotted vertical lines.

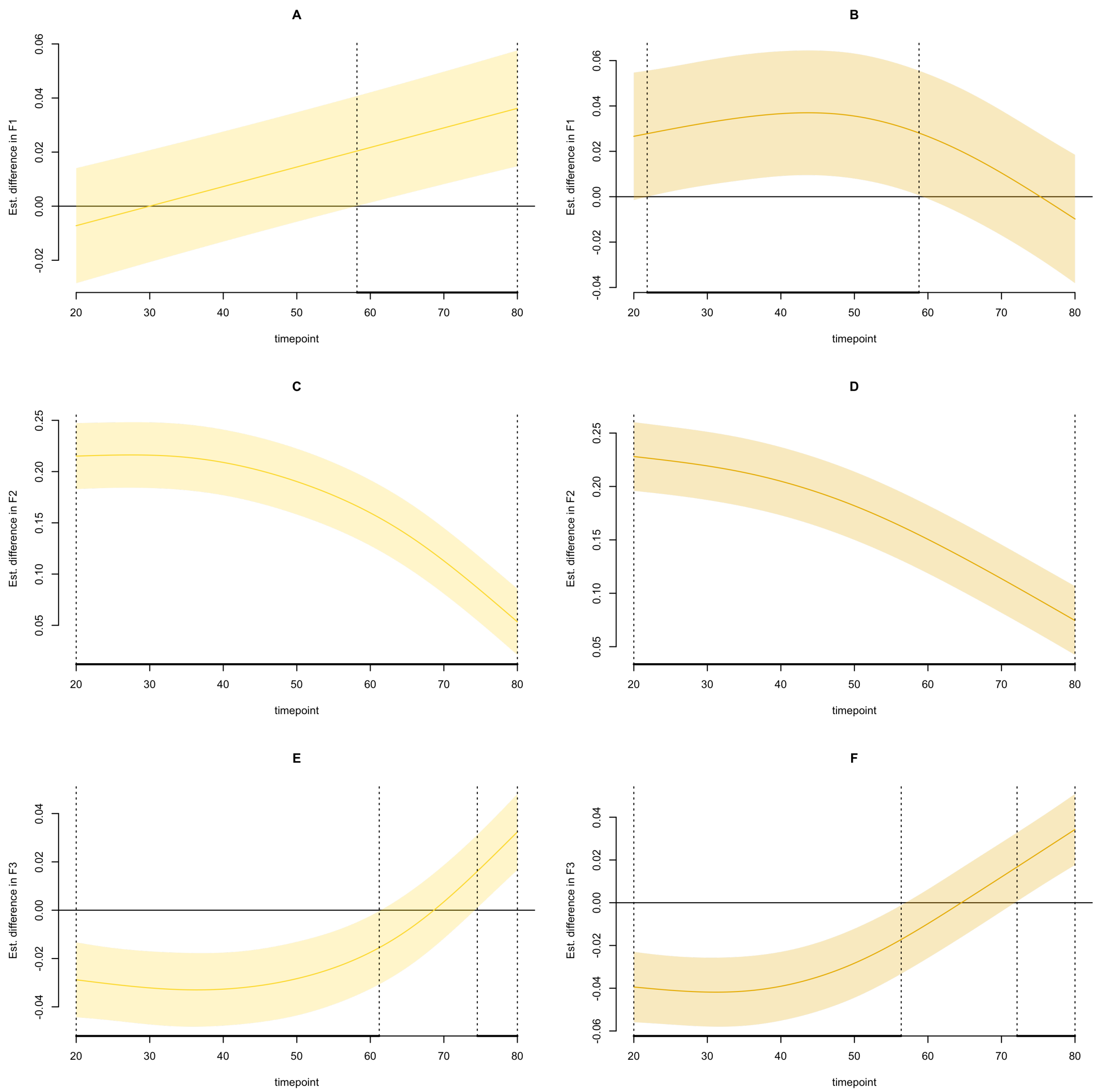

Figure 13. Fitted smooths of GAMM for predicting F1 (**upper row**), F2 (**mid row**), F3 (**bottom row**) and 95% confidence intervals for the [øː] - [œː] contrast (**left**), and the [ø] - [œ] contrast (**right**). Differences significantly different from 0 are marked by black dotted vertical lines.

Finally, an effect of gender was found in two contrasts predicting F1: [iː] - [yː] ($\hat{\beta} = -.075, SE = .026, p < .004$), and [iː] - [ʉː] ($\hat{\beta} = -.06, SE = .024, p < .01$), indicating that the female talkers on average produced these contrasts with higher F1 values than the male talkers. This suggests that the normalization approach likely reduced some talker-specificity related to anatomical differences, but not all. An alternative explanation is that the overall difference in height for these vowels when formant dynamics is considered could be attributed to a real pronunciation difference, driven by sociolinguistic factors (for research on young female talkers driving language change, see e.g., Boberg, 2019; Labov, 1990; for a review, see e.g., Woods, 1997).

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 some contrasts, the vowels overlapped only in parts of the segment, as indicated by the gaps in significant differences in Figures 10**A-B-D-F**, 11**A-C-D-E**, 12**A-C-E**, 13**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. The results further indicated that vowel differences in dynamics were driven by both constant as well as non-linear differences in most cases. The contrasts for which no non-linear differences were found were [ɛ] - [æ] along F3 ($\hat{\beta} = .0099, SE = .005, p > .073$; $EDF = 1, F = .88, p > .34$), [iː] - [yː] along F1 ($EDF = 1.8, F = 1.06, p > .4$) and F2 ($EDF = 1.59, F = 1.0, p > .43$), and [ɪ] - [ʏ] along F2 ($EDF = 1, F = 3.14, p > .076$), which would seem to suggest their similarity in formant movements across the segment.

3.2.2.2 Formant dynamics in long-short vowel pairs The second set of GAMMs modeled the effect of category on F1, F2, and F3 for all long and short vowel pairs. Summary tables and visualizations of these GAMMs are included in the SI (Section S1.8.1).

There was a treatment effect of vowel on all spectral cues for all vowel pairs, driven by both constant (for F1, all $ps < .0067$; for F2, all $ps < .002$, for F3, all $ps < .022$) and non-linear differences (for F1, all $ps < .0004$; for F2, all $ps < .016$, for F3, all $ps < .05$), except for [ɛː] - [ɛ] ($\hat{\beta} = .00003, SE = .008, p > .96$) and [ɑː] - [a] ($\hat{\beta} = -.0043, SE = .01, p > .67$; $EDF = 2.09, F = 1.67, p > .19$) predicting F3. This would thus seem to suggest that F3 does not reliably distinguish between long-short vowels in these pairs when dynamics is considered. Differences in dynamics were overall driven by both constant and non-linear differences. However, some vowel pairs displayed insignificant smooth differences between vowels within pairs, suggesting similarity in formant movements across the segment (for the [uː] - [ʊ] contrast predicting F2, the [æː] - [æ] contrast predicting F1 and F3, and the [œː] - [œ] contrast predicting F3, all $ps > .06$). With the exception of the [yː] - [ʏ] vowel pair, all rounded vowels displayed lower F3 values in their long allophones, presumably indicating greater lip-rounding as caused by more vigorous activity of the lips (e.g., Hadding, Hirose, & Harris, 1976; Stålhammar, Karlsson, & Fant, 1973). Again, significant gender differences were found in height for some high vowels (for the [iː] - [ɪ], [yː] - [ʏ] and [oː] - [ɔ] contrasts predicting F1, all $ps < .0086$), but also in backness (F2) for the [uː] - [ʊ] contrast ($p < .0185$).

The results suggest that when formant dynamics are considered, *all* long-short Central Swedish vowel pairs differ in spectral cues. Among the pairs that displayed smaller, albeit statistically significant, differences in spectral cues is the [iː] - [ɪ] pair predicting F1, possibly indicating a tendency for stronger duration dependency, in line with previous perceptual work (Behne et al., 1997). Overall, larger effects were found for F2 which would support the hypothesis that F2 carries more of the durational variation in quantity contrasts in Swedish. Non-significant differences within pairs were found only for GAMMs predicting F3, which highlights the primary importance of F1-F2 as spectral cues to quantity contrasts under the assumption of formant dynamics.

3.3 Results summary

The results from the static analysis suggest that F1, F2, F3 and duration all contribute to the distinction of Central Swedish vowels. While F1 and F2 are or primary importance for most contrasts, F3 contributes to increasing the separability between neighboring vowels differing in lip-rounding. The static analysis further suggested that while the F1-F2-duration cue combination maximized separability for long-short vowel pairs, the inclusion of F3 also increased separability relative to F1-F2. Importantly, including F3 increased separability more than duration did for the [yː] - [ʏ] contrast.

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, the GAMM fit to the same contrast found no significant differences between the two vowels predicting F3, suggesting less distinguishability under the assumption of formant dynamics. 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.

The resulting phonetic characteristics of the long and short Central Swedish vowels

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]		

presented here, 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).[13] 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 [æː], 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 [ɵ] 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

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

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.

4 General discussion

The purpose of this paper 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 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). The study further 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. Next, the main findings are discussed, alongside methodological considerations and future directions.

Beginning with the static analysis, the results seem to suggest that the most important cues to vowel identity are F1, F2 and duration, which replicates previous work (e.g., Kuronen, 2000; Lindblom, 1963). Including F3 increased separability for all rounding contrasts which suggests that F3 adds information for rounding vs. unrounded neighboring contrasts, given the way category separability is calculated here. The category separability index for the long-short vowel pairs suggested that the F1-F2-duration cue combination maximized separability between vowels within pairs, highlighting the role of both spectral and temporal cues for quantity contrasts and that *all* long-short vowel pairs display spectral differences. The results further indicated that F2 seemed to carry more of the spectral variation in quantity contrasts (Kuronen, 2000; Lindblom, 1963). Of note, however, one long-short vowel pair, [yː] - [ʏ], achieved higher separability with F3-inclusion over Duration-inclusion, despite the fact that the two vowels clearly display systematic differences in duration similar to all other pairs (Figure 5). Across talkers, [ʏ] is clearly more front than [yː] (mean F2 for [ʏ] = 2201 Hz, SD=226; mean F2 for [yː] = 1939 Hz, SD=187), yet F3 is overall higher for [yː] (mean F3 for [yː] = 3058 Hz, SD=259; mean F3 for [ʏ] = 2625 Hz, SD=226). Under the assumption of a lower F3 in high front rounded vowels being indicative of increased lip-rounding (compare, mean F3 for [iː] = 3054 Hz, SD=313; mean F3 for [ɪ] = 2924 Hz, SD=237), this would seem to suggest more rounding in [ʏ] relative to [yː].

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 [iː] - [yː] - [ʉː] - [eː] 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 fit to the long-short vowel pairs indicated that within all pairs, the two categories were significantly different in predicting the spectral cues in all cue evaluations, with the exception of [ɛː]-[ɛ] and [ɑː]-[a] predicting F3. This highlights the informativity carried by formant dynamics for 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. For the [iː] - [yː] contrast, including F3 under static measures either did not appear to change separability (visualizations), or increased the category separability index relative to F1-F2. However, the F1-F2-F3 index for [iː] - [yː] displayed substantial between-talker variability, and was comparatively low. These results overall 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 some of these talkers. A merger might be driven by relaxation of lip-rounding, as supported by the higher F3 values found for [yː].

The vowel space as summarized in Table 2 suggests shifts in some vowels compared to previous characterizations of Central Swedish. For instance, the [ɛː] has lowered substantially compared to earlier mappings of the space (e.g., Engstrand, 1999; Fant et al., 1969; Kuronen, 2000). This lowering was anticipated by Riad (2014) and supported in Leinonen (2010), Gross et al. (2016), and Pelzer and Boersma (2019). Another important shift concerns the fronting of [ɪ] and [ʏ] and the centralization of [iː] and [yː], where [ɪ] and [ʏ] appear to maintain their positions as high front vowels, while [iː] and [yː] have centralized to the mid-center part of the space (see also, Pelzer & Boersma, 2019).[14] This conflicts with previous work on Swedish as well as other languages demonstrating that short vowels are generally less peripheral and more centralized than their long counterparts (e.g., Clopper et al., 2005; Hillenbrand et al., 1995). Besides the already mentioned hypothesis of a possible [iː] - [yː] merger due to relaxation of lip-rounding in [yː], an alternative, yet related, hypothesis is that both, or either, vowels are produced as damped versions, as previously found for talkers of Stockholm Swedish (e.g., Kotsinas, 1994; Schötz et al., 2011). 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. Impressionistic listening by the author did support the presence of a final consonantal glide, similar to [ɨː], among the majority of talkers in SwehVd, both male and female. The strength and scope of [ɨː] varied across talkers, from a relatively strong [ɨː] (18 talkers), to a weaker voiced fricative offglide buzzing [z] (14 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

[14] 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.

consonant glide or buzz (12 talkers), seemed to have overall less retracted [iː] and [yː], hence supporting the hypothesized link between centralization and consonantal offglide.

4.1 Methodological considerations and future directions

Two sets of methodological considerations for the present work not mentioned elsewhere, deserve further discussion. The first set concerns the measuring and evaluation of two of the acoustic cues, F0 and vowel duration. F0 is not considered an important cue to vowel identity in itself and therefore often not included in previous work (e.g., Fant et al., 1969; Gross et al., 2016; Leinonen, 2010; Pelzer & Boersma, 2019; Wenner, 2010). However, it is known to vary across languages and dialects (e.g., Henton, 2005; Jacewicz & Fox, 2018; Johnson, 2005; Leung et al., 2016; Mennen et al., 2012; Weirich et al., 2019), and has been shown to have strong indirect effects on vowel categorization (Barreda & Nearey, 2012; see also work on vowel-intrinsic F0, e.g., Whalen & Levitt, 1995; and the hypothesized use of F0 for tense-lax distinctions in German, Pape & Mooshammer, 2006). For the above reasons, analysis of F0 was included in this study but limited to reports of mean F0 across the segment for static visualizations. Figure 3 confirmed that F0 carried the least information about vowel identity of all cues included in this material. However, given evidence of pitch contours influencing perceived duration and prominence of vowels (e.g., Gussenhoven & Zhou, 2013), one could claim that a more comprehensive acoustic representation of vowels should have included F0 in the dynamic analysis. It is not unreasonable to assume a role for F0 in crowded systems such as the Central Swedish vowel inventory with category overlap and potential mergers (c.f., research on adaptive dispersion, e.g., Liljencrants & Lindblom, 1972; Lindblom, 1998). For instance, talkers might combine tone distinction (F0) with voice quality (e.g., creak, or buzzing, as in the damped [ɨː]) to increase distinctions between neighboring contrasts (for a review, see e.g., Davidson, 2021).

For duration, the temporal analysis of the effect of duration on long-short vowel pairs could have been supplemented with measures of consonant ratios, following, e.g., Pelzer and Boersma (2019), and Schaeffler (2005). Since the SwehVd database is publicly available in an online repository (https://osf.io/ruxnb/), both the question of consonant ratios and the role of F0 can be addressed in future studies.

The second methodological consideration concerns methods used for assessing category distinguishability. Since the GAMMs were fit to each cue separately, they allowed for separate assessments of the relative weight of each cue, compared to the separability index where the by-cue contributions could 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* distinguishability, even if it is reasonable to assume that reduced overlap between tokens of neighboring categories would increase intelligibility (e.g., Bradlow, 1995; Wright et al., 2004). For this, one could fit other types of models, such as a multinomial logistic regression model predicting vowel category from cue combinations, or perceptual models assessing the predicted consequences for perception. In separate work conducted in parallel with this study, we have pursued a similar approach, evaluating the predicted consequences of F3-inclusion for high-front vowel contrasts in Swedish using a perceptual model based on Bayesian inference, ideal observers (Persson & Jaeger, 2024). The results of including F3 qualitatively replicated the results presented here for the investigated vowels in that it overall improved the predicted recognition accuracy. Compared to measures of category separability, the ideal observers allowed for an assessment of the effect of F3-inclusion on category confusability among all vowels considered. This analysis suggested that while F3-inclusion overall decreased category confusability, especially for [yː] and [ʏ], it *increased* the probability of confusing [iː] with [yː].[15]

Models predicting perception from production data can further inform 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.

Another avenue for future research to explore, are the preliminary hypotheses mapped out in this section concerning vowel change. Since this study's primary focus is mapping the acoustics of modern-day Central Swedish vowels, systematic investigations of the underlying reasons to the potential shifts in the vowel space compared to previous work, are left for future studies. These hypotheses, and others, could be investigated in perceptual studies or by using systematic listening by trained phoneticians, further

[15] In comparison to other model-based approaches such as logistic regression and linear discriminant analysis, Bayesian ideal observers have the advantage of reducing the number of degrees of freedom in the fit from production data to predicting perception, which substantially reduces the risk of over-fitting to the data (see discussion in e.g., Persson & Jaeger, 2023).

validated through measures of inter-rater reliability (for a review, see Cucchiarini, 1995; Gross & Forsberg, 2020; Kuronen, 2000; Pelzer & Boersma, 2019). Evaluations by trained phoneticians can provide assessments of both auditory distinguishability between categories and the correspondence between assigned IPA label and auditory impression. Perceptual studies with naïve listeners can complement evaluations by phoneticians, as training can differ across phoneticians and introduce biases (for reviews, see e.g., Heselwood & Howard, 2008; Kerswill & Wright, 1990; Stemberger & Bernhardt, 2020). Relatedly, the extent to which individual talkers are driving these changes could be investigated by assessing the amount of cross-talker variability in SwehVd. This study suggests that even when keeping the background variables age and region of origin constant and normalizing for talker-specificity in physiology, there are still differences in the phonetic realization of some vowels that could be related to sociolinguistic differences. Further insights into these individual differences can be gained by studying the SwehVd materials on a talker-specific level.

5 Conclusions

The present study has reported on the acoustic properties of Central Swedish vowels. 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 relaxation of lip-rounding in [yː], the lowering of [ɛː], and the potential role of additional cues beyond those investigated here, 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, Peter F., & William F. Katz. (2005). Synthesis fidelity and time-varying spectral change in vowels. *The Journal of the Acoustical Society of America*, *117*(2), 886–895. https://doi.org/10.1121/1.1852549

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

Barreda, Santiago. (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, Santiago, & Terrance M. Nearey. (2012). The direct and indirect roles of fundamental frequency in vowel perception. *The Journal of the Acoustical Society of America*, *131*(1), 466–477. https://doi.org/10.1121/1.3662068

Barreda, Santiago, & Terrance M. Nearey. (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, Dawn M., Peter E. Czigler, & Kirk P. H. Sullivan. (1997). Swedish Quantity and Quality: A Traditional Issue Revisited. *Reports from the Department of Phonetics, Umeå University*, *4*, 81–83.

Björsten, Sven, & Olle Engstrand. (1999). Swedish "damped" /i/ and /y/: Experimental and typological observations. *Proceedings of the 14th International Congress of Phonetic Sciences*, 1957–1960. San Fransisco.

Bleckert, Lars. (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.

Boberg, Charles. (2019). A closer look at the short front vowel shift in Canada. *Journal of English Linguistics*, *47*(2), 91–119. https://doi.org/10.1177/0075424219831353

Boersma, Paul, & David Weenink. (1992-2022). *Praat: Doing phonetics by computer [Computer program].*

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

Bruce, Gösta. (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

Bruce, Gösta. (2010). *Vår fonetiska geografi : Om svenskans accenter, melodi och uttal [Our phonetic geography: about the accents, melody and pronunciation of Swedish]* (första upplagan). Lund: Studentlitteratur.

Chesworth, Janine, Kim Coté, Colleen Shaw, Sandra Williams, & Megan Hodge. (2003). Effect of phonetic context on women's vowel area. *Canadian Acoustics/Acoustique Canadienne*, *31*, 20–21.

Chuang, Yu-Ying, Janice Fon, Ioannis Papakyritsis, & Harald Baayen. (2021). Analyzing phonetic data with generalized additive mixed models. In *Manual of clinical phonetics* (pp. 108–138). Routledge.

Clopper, Cynthia G., David B. Pisoni, & Kenneth De Jong. (2005). Acoustic characteristics of the vowel systems of six regional varieties of American English. *The Journal of the Acoustical Society of America*, *118*(3), 1661–1676.

Cucchiarini, Catia. (1995). Assessing transcription agreement: Methodological aspects. *Clinical Linguistics and Phonetics*, *10*(2), 131–155. https://doi.org/10.3109/02699209608985167 0269-9206

Davidson, Lisa. (2021). The versatility of creaky phonation: Segmental, prosodic, and sociolinguistic uses in the world's languages. *Wiley Interdisciplinary Reviews: Cognitive Science*, *12*(3), e1547.

Eklund, Ingegerd, & Hartmut Traunmüller. (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, Claes-Christian. (1964). *Phonologic studies of quantity in Swedish: Based on material from Stockholm speakers*. Uppsala: Almqwist & Wiksell.

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

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

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

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

Engstrand, Olle, Sven Björsten, Björn Lindblom, Gösta Bruce, & Anders Eriksson. (2000). Hur udda är Viby-i? Experimentella och typologiska observationer [How peculiar is Viby-i? Experimental and typological observations]. *Folkmålsstudier*, *39*, 83–95.

Fabricius, Anne, Dominic Watt, & Daniel Ezra Johnson. (2009). A comparison of three speaker-intrinsic vowel formant frequency normalization algorithms for sociophonetics.

Language Variation and Change, *21*(3), 413–435. https://doi.org/10.1017/S0954394509990160

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

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

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

Flynn, Nicholas, & Paul Foulkes. (2011). Comparing vowel formant normalization methods. *Proceedings of ICPhS XVII, Hong Kong*, (August), 683–686.

Fujimura, Osamu. (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

Gross, Johan. (2018). *Mapping vowels: Variation and change in the speech of Gothenburg adolescents*. Gothenburg: University of Gothenburg.

Gross, Johan, Sally Boyd, Therese Leinonen, & James A. Walker. (2016). A tale of two cities (and one vowel): Sociolinguistic variation in swedish. *Language Variation and Change*, *28*(2), 225–247.

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

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

Hadding, Kerstin, Hajime Hirose, & Katherine S. Harris. (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, Kerstin, & Arthur S. Abramson. (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., & L. Norman. (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).

Henton, Caroline 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

Heselwood, Barry, & Sara Howard. (2008). Clinical phonetic transcription. *The Handbook of Clinical Linguistics*, 381–399.

Hillenbrand, James M., Michael J. Clark, & Terrance M. Nearey. (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, James M., Laura A. Getty, Michael J. Clark, & Kimberlee Wheeler. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America*, *97*(5), 3099–3111.

Hillenbrand, James M., & Terrance M. Nearey. (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

Holbrook, Anthony, & Grant Fairbanks. (1962). Diphthong formants and their movements. *Journal of Speech and Hearing Research*, *5*(1), 38–58.

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

Jacewicz, Ewa, Robert Allen Fox, & Joseph Salmons. (2011). Regional dialect variation in the vowel systems of typically developing children. *Journal of Speech, Language, and*

Hearing Research, *54*(2), 448–470. https://doi.org/10.1044/1092-4388(2010/10-0161)

Jenkins, James J., Winifred Strange, & Salvatore Miranda. (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, Keith. (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, Keith, & Matthias J. Sjerps. (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, Martin. (1948). Acoustic Phonetics. *Language*, *24*(2), 5–136. https://doi.org/10.2307/522229

Kent, Raymond D., & Houri K. Vorperian. (2018). Static measurements of vowel formant frequencies and bandwidths: A review. *Journal of Communication Disorders*, *74*, 74–97. https://doi.org/https://doi.org/10.1016/j.jcomdis.2018.05.004

Kerswill, Paul, & Susan Wright. (1990). The validity of phonetic transcription: Limitations of a sociolinguistic research tool. *Language Variation and Change*, *2*(3), 255–275. https://doi.org/10.1017/S0954394500000363

Kotsinas, Ulla-Britt. (1994). *Ungdomsspråk [Youth language]*. Uppsala: Hallgren & Fallgren.

Kuronen, Mikko. (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, William. (1990). The intersection of sex and social class in the course of linguistic change. *Language Variation and Change*, *2*(2), 205–254. https://doi.org/10.1017/S0954394500000338

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

Labov, William. (2010). *Principles of linguistic change. 2: Social factors* (repr). Chichester: Wiley-Blackwell.

Labov, William, Sharon Ash, & Charles Boberg. (2005). *The atlas of North American English: Phonetics, phonology, and sound change.* Berlin • New York: De Gruyter Mouton. https://doi.org/doi:10.1515/9783110167467

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

Lehiste, Ilse, & Gordon E. Peterson. (1961). Transitions, glides, and diphthongs. *The Journal of The Acoustical Society of America*, *33*(3), 268–277.

Leinonen, Kari, Pitkänen Antti J., & Veijo V. Vihanta. (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, Therese. (2010). *An Acoustic Analysis of Vowel Pronunciation in Swedish Dialects.* Groningen: University of Groningen.

Leung, Keith KW, Allard Jongman, Yue Wang, & Joan A. Sereno. (2016). Acoustic characteristics of clearly spoken English tense and lax vowels. *The Journal of the Acoustical Society of America*, *140*(1), 45–58.

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

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

Lindblom, Björn. (1998). Systemic constraints and adaptive change in the formation of sound structure. In H. J. R., S.-K. M., & K. C. (Eds.), *Approaches to the evolution of language: Social and cognitive bases* (pp. 242–264). Cambridge University Press

Cambridge, United Kingdom.

McAllister, Robert, James Lubker, & Johann Carlson. (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, Ineke, Felix Schaeffler, & Gerard Docherty. (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

Munson, Benjamin, & Nancy Pearl Solomon. (2004). The effect of phonological neighborhood density on vowel articulation. *Journal of Speech, Language, and Hearing Research*, *47*(5), 1048–1058. https://doi.org/10.1044/1092-4388(2004/078)

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

Nearey, Terrance M., & Peter F. Assmann. (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

Nilsson, Jenny, Lena Wenner, Therese Leinonen, & Eva Thorselius. (2021). New and old social meanings in urban and rural Sweden. *Urban Matters*, 179.

Nycz, Jennifer, & Lauren Hall-Lew. (2013). Best practices in measuring vowel merger. *Proceedings of Meetings on Acoustics*, *20*. AIP Publishing.

Pape, Daniel, & Christine Mooshammer. (2006). Intrinsic F0 differences for German tense and lax vowels. *Proceedings of the 7th International Seminar on Speech Production in Ubatuba, Brazil, December 13th to 15th*, 271–278.

Pelzer, Joppe Anna, & Paul Boersma. (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, Anna, Santiago Barreda, & T. Florian Jaeger. (2024). *Comparing accounts of formant normalization against US English listeners' vowel perception*. Manuscript;

Stockholm University.

Persson, Anna, & T. Florian Jaeger. (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

Persson, Anna, & T. Florian Jaeger. (2024). Measuring the informativity of F3 for rounded and unrounded high-front vowels in Central Swedish. *Proceedings from FONETIK 2024, Stockholm, June 3–5, 2024*, 13–18. Stockholm University. https://doi.org/10.5281/zenodo.11396050

Peterson, Gordon 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, Gordon E., & Harold L. Barney. (1952). Control methods used in a study of the vowels. *The 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, Margaret E. L., & Joseph A. Stanley. (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, Tomas. (2014). *The phonology of Swedish.* Oxford: Oxford University Press.

Robb, Michael P., & Yang Chen. (2009). Is /h/ phonetically neutral? *Clinical Linguistics & Phonetics*, *23*(11), 842–855. https://doi.org/10.3109/02699200903247896

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

Scarborough, Rebecca. (2012). Lexical similarity and speech production: Neighborhoods for nonwords. *Lingua*, *122*(2), 164–176. https://doi.org/https://doi.org/10.1016/j.lingua.2011.06.006

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

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

Schertz, Jessamyn, & Emily J. Clare. (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, Susanne, Johan Frid, & Anders Löfqvist. (2011). Exotic vowels in Swedish – an articulographic and acoustic pilot study of /iː/. *Proceedings of the 17th International Congress of Phonetic Sciences*. Hong Kong.

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

Sóskuthy, Márton. (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, Márton, Paul Foulkes, Vincent Hughes, & Bill Haddican. (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, Carl Ivar. (1965). 'Mötet uppnas på sundag'[The meeting uppnas on sundag]. *Språkvård*, *3*(3-8), 1–15.

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

Stemberger, Joseph Paul, & Barbara May Bernhardt. (2020). Phonetic transcription for speech-language pathology in the 21st century. *Folia Phoniatrica Et Logopaedica*, *72*(2), 75–83.

Stevens, Kenneth N., & Arthur S. House. (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

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

Strange, Winifred. (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, Eva. (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, Ann 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, Catherine I., & Jonathan Harrington. (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, Andrew, Noah Nelson, & Rebecca Sharp. (2018). The phonetic specificity of contrastive hyperarticulation in natural speech. *Journal of Memory and Language*, *100*, 61–88.

Weirich, Melanie, Adrian P. Simpson, Jasmine Öjbro, & Christine Ericsdotter Nordgren. (2019). The phonetics of gender in Swedish and German. *Fonetik 2019, Stockholm, Sweden, 10-12 June, 2019*, 49–53.

Wenner, Lena. (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.

Whalen, D. H., & Andrea G. Levitt. (1995). The universality of intrinsic F0 of vowels. *Journal of Phonetics*, *23*(3), 349–366.

https://doi.org/https://doi.org/10.1016/S0095-4470(95)80165-0

Wieling, Martijn. (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.

Woods, Nicola J. (1997). The formation and development of New Zealand English: Interaction of gender-related variation and linguistic change. *Journal of Sociolinguistics*, *1*(1), 95–125. https://doi.org/https://doi.org/10.1111/1467-9481.00005

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

Xie, Xin, & T. Florian Jaeger. (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

Yang, Byunggon. (2019). A comparison of normalized formant trajectories of English vowels produced by American men and women. *Phonetics and Speech Sciences*, *11*(1), 1–8.

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