LUND UNIVERSITY

JOINT FACULTIES OF HUMANITIES AND THEOLOGY

Formant Dynamics and Word Accents in Central Swedish

Vide Nordin Nobuoka
Centre for Languages and Literature (SOL)
Bachelor's Thesis in Phonetics
Supervisor: Mechtild Tronnier
Spring 2026

Abstract:

Studies of intrinsic F0 has mostly investigated static vocalic aspects, while the interactions between F0 and formant movements is a largely unexplored field. This thesis explores the interactions between movements in F0 and formants in Central Swedish, a dialect with word accent pitch contours and diphthongized vowels. Inter- actions between creaky voice and formant movement, and creaky voice and word accents are also studied. A reading elicitation task was used as material for the present study. The results showed that F0 movement may affect the degree of diphthongization; that F0 movements and diphthongized offglides tended to cooccur; and that accent II may facilitates greater formant movements in certain vowels in central Swedish. The results lended nothing conclusive about the interactions between creaky voice and formant movements, but showed that creaky voice is more likely to appear on accent I words. Insertions of labial approximants during /ʉː/, /uː/, and /oː/ were found in the studied material.

Acknowledgements:

First, I want to thank my supervisor Mechtild Tronnier for supporting me by helping me overcome hurdles in writing this paper and being excited for this project. My gratefulness extends, as always, to my wonderful partner for your endless support and helping me be more organized. I also want to thank my course mates, Anna, Anton, Göran, Helle, Helmer, and Jonathan for inspiring me with their own works. The author gracefully acknowledges Lund University Humanities Lab, and I want to thank Peter for your patience and expertise in the recording studio. Finally, thank you to all the speakers who took time out of your days to make this study possible.

Table of contents:

1. Introduction:

This study explores how vowel formant dynamics interact with fundamental frequency, creaky voice, and Swedish word accents.

The interactions between segmental and prosodic features is a part of phonetics still underexplored. Acoustic phonetics as a science has deepened our insight into the physics of language. In interacting with other linguistic fields, acoustics has helped develop speech pathology, in addition to speech technologies. Further understanding of the acoustics of speech enables further communicatory and linguistic discoveries. The complexity of speech is still rife with unanswered questions, and in order to answer those questions, we need to look at specific cases. This thesis takes a closer look at the case of the interaction between fundamental frequency (F0) and the dynamics of vowels. The rapidity of speech gives rise to coarticulations, where multiple phonemes affect a single segment (Gick et al., 2013). This is apparent in the acoustics of vowels, where the alternations between consonant and vowel manifest as formant transitions (Baart, 2010, p. 72-74). F0 is the acoustic correlate to the auditory phonetic feature pitch, and is measured in hertz (Hz). The degree of tension in the vocal folds determines the rate of vibration, and in turn the pitch of a voiced segment (Crystal, 2008). Since the vocal tract includes the articulator of tones, logically they should affect each other. One such effect is the intrinsic pitch of vowels, where the height of a vowel affects its perceived pitch (Ewan, 1975). This effect is, however, mostly explored concerning monophthongs, and diphthong interactions with intrinsic pitch and F0 are only recently getting more attention (Niebuhr, 2004; Li & Al-Tamimi, 2026). Diphthongs are vowels that have two targets (Lehiste & Peterson, 1961). Vowels, and diphthongs in particular, are interesting objects of study because their inherent dynamics can be compared to the dynamics of F0.

Creaky voice is a voice quality articulated with short and thick vocal folds, often accompanying low pitch (Esling et al., 2019), and has been found to coincide with the Swedish pitch accents differentially (Heldner et al. 2021; Hjortdal, 2022).

The variety of Swedish studied here is a type of dialect spoken in and around Stockholm, the capital city of Sweden, sometimes called Central Swedish (see section 2.5). Though Central Swedish lacks phonological diphthongs (Eliasson, 2022), diphthongization of vowels has been attested in Central Swedish and surrounding dialects (Bruce, 2010; Kuronen, 2000; Pelzer & Boersma, 2019). The phonology of Central Swedish also contains pitch accents, called accent I and accent II, which differ in their tonal contours (Bruce, 1977; Riad, 1998). These pitch accents are sometimes called word accents (Bruce, 2005; Heldner et al., 2021), though word accent is a broader category (Crystal, 2008). For the purposes of this thesis, the two terms are interchangeable.

1.1 Research questions

1. Do the dynamics of F0 affect the dynamics of vowel formants?
    1.1. Do the movements of a dynamic F0 coincide with formant movements in diphthongized vowels?
2. Does the occurrence of creaky voice coincide with formant movements in diphthongized vowels?
3. Do the word accents in Central Swedish differ in their facilitation of vowel dynamics and dipthongization in the first vowel of two-syllable words with stress on the first syllable?
4. Do the word accents in Central Swedish differ in their facilitation of creaky voice during the first vowel of two-syllable words with stress on the first syllable?

2. Background

A diphthong is a vocalic segment during which there is a perceptually noticeable change in vowel quality (Crystal, 2008, p. 146). Diphthongization is a perceptually observable shift of the quality of a vowel during its articulation, presupposing an underlying phonological monophthong (Kuronen, 2000). Kuronen (2000) illustrates that diphthongization of monophthongs in Finnish, a language with diphthong phonemes, is prohibited. E.g. pronouncing /e:/ as [ie] would trigger perception of the diphthong /ie/. Diphthongs can be divided into rising/falling and opening/closing. The rising/falling distinction concerns wether the first or second component of the diphthong is a glide; falling diphthongs have final glides while rising diphthongs begin with a glide (Crystal, 2008). The opening/closing distinction concerns the relation between the two target vowel qualities, and wether the second component has a target that is more open or more closed than the first component (Niebuhr, 2004).

2.1 Interactions between diphthong shape and prosodic features

Different vowels have different intrinsic pitch. What this means is that there is a correlation between vowel quality and F0, where vowel height is positively correlated with F0. One hypothesis about where this correlation comes from, which states that the degree of tongue pull on the larynx affects the degree of vocal cord tension, has some empirical evidence: when articulating high vowels, the body of the tongue is pushed up towards the palate, pulling the larynx up and stretching the vocal folds (Ohala & Eurkel, 1976). This effect is seen crosslinguistically, but to different degrees in different languages. The cause of this difference cannot be explained by a single factor, though minor effects might include vowel inventory size and existence of tonal features; Intrinsic pitch may have a small enhancing effect in the differentiation between vowels that are similar, and it may have a small suppressing effect when it comes to differentiation between tones (Ting et al., 2025).

Li & Al-Tamimi (2024) investigated the effects of F0 on diphthong relizations in Standard Mandarin /ai/ and /au/. They found that different tone contours affect vowel height, as F0 generally correlates negatively with F1. A diphthong with H lexical tone was pronounced more closed, one with a L tone was pronounced more open, one with a rising tone was pronounced more diphthongized, and a diphthong with a falling lexical tone was pronounced more monophthongized. It has also been found that the falling tone in Standard Mandarin increases the duration of diphthong segments in both men and women, and that the L tone increases the duration of diphthongs produced by women, but not men (Li et al., 2023).

If the findings of this study were to follow what was found by Li and Al-Tamimi (2024), then vowels in accent II without focus (H*L) can be hypothesized to be pronounced more opening, i.e. opening diphthongs more diphthongized and closing diphthongs more monophthongized. Conversely, vowels in accent I with focus (L*H) can be hypothesized to be pronounced more closing.

Another side to the relationship between F0 and diphthongs is the effects on perception. Niebuhr (2004) tested whether or not the closing and opening diphthong articulations found in German correlated with a rising and falling pitch perception. He hypothesized that a closing diphthong would support the perception of a falling pitch contour, and that a opening diphthong would support the perception of a rising pitch contour. Additionally, these effects were hypothesized to be more pronounced the more back and rounded the close component of the diphthong is. The results of that study showed that the opening diphthongs of [iɐ] and [uɐ] have intrinsic rising pitch contours of around +0,25 % and +1,38 %, respectively. Notably, these two vowels are similar to the diphthongized /eː/ and /oː/ of Central Swedish.

Based on the findings of Niebuhr (2004), another hypothesis can be made. In a falling pitch contour, opening diphthongized vowels such as /eː/ and /oː/, may facilitate a greater fall in pitch, compared to monophthongs or closing diphthongized vowels, since their intrinsic F0 triggers the perception of a rising pitch in flat F0 contours.

2.2 Temporal aspects of diphthongs

In his book Diphthong Dynamics, Peeters (1991) described a hypothesis of diphthong articulation where the salient identifier of a diphthong is defined, not just by onset and offset formant values, but the durational relationship of the diphthong pattern components of steady-state and glide. Different languages are proposed to have different temporal patterns for the same diphthongs, i.e. the English /au/ and the German /au/ differ temporally. Two diphthong pattern types: tripartite (steady-state/glide/steady-state) and bipartite (steady-state/offglide) were found in the study. English was found to have a significant bi- partite pattern, with relatively long onset steady-states, while German was found to have a tripartite pattern with relatively short onset- and long offset steady-states. In order for a vowel to be perceived as a diphthong (instead of as a sequence of two monophthongs), it was found that the glide needs to be at least 80 ms long. The effect that F0 had on integration was inter- esting, as a dynamic F0 caused integration for German diphthongs, but caused segregation for the other languages. The duration of the final component of a vowel was found to be espe- cially significant in vowel identification. Peeters (1991) also proposes that a universal vocalic temporal integration constant of around 80 ms may exist.

2.3 Creaky voice

Creaky voice is a phonation type (also called voice quality) where the thyroarytenoid muscle draws the arytenoid cartilage towards the thyroid, causing the vocal folds to shorten and relax (Esling et al., 2019, p. 63). Creaky voice occupies the range below 100 Hz (Esling et al., 2019, p. 63-66), as opposed to modal voice, whose pitch range averages around 120-220 Hz, depending on speaker (Esling et al., 2019, p. 46). Creaky voice is also called ‘vocal fry’ (Crystal, 2008, p. 121). Creaky voice is often used as cues for signalling the ends of phrases (Garellek & Keating, 2015).

Heldner et al. (2021) investigated Central Swedish word accents, showing that they coincide with the tonal word accent distinction. The material for that study were words produced in citation form. In their study, creaky voice (though rare) was more probable in vowels with the falling L tone associated with the second syllable in accent I words. Hjortdal (2022) investigated South Swedish word accents using non-focally produced words as the material. The results in her study suggested that creaky voice was employed more often in accent I words, and that creaky voice may be cue for word accents. The model by Lindqvist-Gauffin (1972) and Laryngeal Articulator Model (Esling et al., 2019) propose that laryngeal constriction corresponds with a falling pitch, which aligns with the findings of Hjortdal (2022).

2.4 Swedish vowels: monophthongs and diphthongizations

In his description of the Central Swedish vowel space, Riad (2013b) defines 9 phonemic vowel categories: /i/, /y/, /ʉ/, /e/, /ɛ/, /ø/, /a/, /o/, and /u/. Each vowel category has a pair of long vs. short, except for /eː/ and /ɛː/ where their short counterparts have collapsed into [ɛ], that manifest as spectrally different (Riad, 2013b). The total phonemically different vowel qualities are thus 17.

Most of the recent acoustic descriptions of the vowel space of Swedish have focused on certain areas of the space (Persson, 2025). The SwehVd database contains recordings of speakers of Central Swedish with every vowel category in Swedish represented, including the allophones [æ] and [œ] (Persson & Jaeger, 2025). Using the material in SwehVd, Persson (2025) found that not only temporal and static spectral cues, but also dynamic spectral cues distinguish Swedish vowel categories. Also when comparing short and long vowels, the spectral values and movements were salient. She noted that [eː] and [oː] displayed clearer patterns of diphthongization.

Of the long vowels, /aː/ and /ɛː/ have been described as the only monophthongal qualities (Eklund & Traunmüller, 1997), while mid vowels tend to diphthongize opening and high vowels tend to diphthongize closing (Kuronen, 2000). The closing diphthongization of the high vowels has also been described as approximant (Riad, 2013b) or fricative offglide (Persson, 2025). The diphthongization of Central Swedish vowels detailed in Kuronen (2000) is summarized in the following table:

iː	Closing; falling F1, rising F2, rising F3. Frication, strong final. Release phase.
yː	Closing; falling F1, rising F2. Frication. Release phase.
ʉː	Closing; falling F1, falling F2, slightly falling F3. Gradual spectral change. Perceptually subtle.
eː	Distinctly opening; rising F1, falling F2. Formant values similar to [i] until 60 % into the vowel. Clear transitional phase.
ɛː	Distinctly opening; rising F1, falling F2. Steady first half. Idiolectal variation.
øː	Opening; rising F1, falling F2. Gradual spectral change. Idiolectal variation.
aː	Opening; rising F1, rising F2. Rising diphthongization. Perceptually subtle.
oː	Opening; rising F1, rising F2. Steady until 50 % into the vowel. Clear transisitonal phase. First half similar to /uː/; End phase similar to /aː/.
uː	Closing; falling F1, falling F2. Perceptually subtle. First half similar to /oː/

Table 1: Characterization of the diphthongization of Central Swedish vowels (Kuronen, 2000). The vowels with gradual spectral change displayed no clear phases/components. Perceptually subtle vowels were barely auditorily identifiable as diphthongized.

Kuronen (2000) also identifies that the transitional phase of diphthongized /eː/ may be more perceptually salient than a final steady-state phase. The release phase of /iː/ and /yː/ was only found sometimes in prepausal, one-syllable words, and therefore was not seen as a feature of the vowels (Kuronen, 2000, p. 88-89).

2.5 Pitch accents

2.5.1 Prosodic features of Swedish

Swedish pitch accents, called accent I and accent II, are a prosodic feature found in Swedish that distinguish words through F0 contour (Myrberg & Riad, 2015). Different phonological descriptions of these word accents adhere to different definitions when it comes to their markedness (Riad, 2013a, p. 181-182). According to the privative representation, one of the accents is marked and the other is a product of sentence intonation (Riad, 1998; Myrberg & Riad, 2015). According to the equipollent view, the difference between the two word accents can be found in the timing of a basic tonal gesture in relation to the stressed syllable of a word (Bruce, 2005).

Bruce (2005) presents the analysis of a basic L+H+L tonal gesture manifesting as a bitonal word accent (either as L+H or H+L, depending on the dialect type), stating the timing of accent I precedes the timing of accent II, regardless of dialect type. Bruce (2005) further elaborates that word accent pitch contours interact with sentence-level tones such as focus and boundary tones. The basic structure of Swedish focus is characterized as a rise to a maximum that is higher than basic word accent maxima (H^-) (Bruce, 1977). The basic contour of Swedish final boundary tones is a fall from a maximum, where the minimum (L%) is reached at the middle of the ultimate vowel of a sentence (Bruce, 1977). This boundary tone applies to isolated declarative sentences, i.e. not questions and not lists.

Riad (2013a) presents a privative analysis where accent II is the marked category, bearing tone as a part of the lexical representation of certain morphemes, and accent I is an empty category, i.e. the tonal features of accent I are only on the intonational level.

Riad (1998) presents the following overview of the types of dialects, and where they are spoken:

Accent I	Accent II	Focus tone	Boundary tone	Dialect
Ø	H*	LH	L	Stockholm
Ø	L*	HL	H	Malmö
Ø	H*	L	H	Göta
Ø	L*	H	L	Dala
Ø	Ø	H	L	Helsinki

Table 2: A typological overview of the dialect types of scandinavian pitch accents. Adapted from Riad (1998).

2.5.2 The 2A pitch accent type

The dialect type studied here, Central Swedish, is of the 2A type (also known as SVEA), which is described as having a low (L) tone associated with the stressed vowel of accent I words, and a falling (H+L) tone associated with the stressed vowel of accent II words. The L tone in accent I words is proceeded by a high (H) tone which falls sharply before the stressed vowel (Bruce, 2007; Bruce, 1977). These descriptions apply for the basic contour of word accents, and are interrupted by other tonal features such as sentence focus (Bruce, 1977). Focus adds a H shortly after the word accent gesture (Bruce, 2005), resulting in undershoot of the word accent L tones.

In the material gathered for this study, the pitch contours of the target vowels are assumed to fall into the four basic tonal contours. These contours have been reviewed in Myrberg & Riad (2015).

	accent I	accent II
Focus	L*H	H*LH
No focus	HL*	H*L

Table 3: The basic tonal contours found in Central Swedish. Properties of lexical tone (columns) and intonational tone (rows) determine contour shape.

3. Method:

No generative AI has been used in this study.

Participants took part in a reading task in which they read a short story containing words with target vowels in accent I and accent II. The procedure, material, participants, and the pilot study are detailed below.

3.1 Procedure of recording

The participants were recorded in the LARM-studio of Lund University Humanities Lab, at Lund University. The following equipment was used: DPA 4066 omnidirectional headset microphone, Neumann U87 Ai microphone, Neve 5088 sound mixer, Antelope Audio Orion 32+ Gen 3 analog-to-digital converter, and Logic Pro software. For each participant, the elicitation material was presented as a laminated A4 paper (Times New Roman, 15p, 1.5 spaced), placed on a music stand. A Neumann microphone was placed above the music stand, and a headset microphone was worn by the participants. Because the headset microphone would be disturbed by earrings worn by some participants, those participants were asked to remove those earrings during the recording. The participants stood up during the recording. The participants were instructed to speak in a calm tempo, and were informed that they could read the material multiple times if they wanted to. If a certain target word was noticeably slurred or mispronounced, I would ask them to reread the sentence with the target word. Before being asked to read the elicitation material, some small talk was engaged between myself and the participant in order for them to become accustomed to the studio environment and setup. The objective of this was for the elicitation to be more implicit, as a more comfortable participant was assumed to speak more naturally.

3.2 Elicitation material

The elicitation material was designed so that every vowel phoneme is represented twice, once in a stressed syllable with accent I and once in a stressed syllable with accent II. The target words were all two-syllable simplex words. The codae in the stressed syllables were limited to /n/ or /l/ in order to minimize preaspiration and to standardize the final formant transitions of the studied vowels. One pair of target words had the rhyme /umː/, due to the restricted distribution of short [u]. Unlike Persson (2025), I have chosen to exclude the allophonic vowels [æ] and [œ] which are in complementary distribution with [ɛ] and [ø] respectively, as objects in this study. The complete elicitation material can be found in the appendix. In this study, fast speech is detrimental since the object of study are vowel dynamics, which overlap with consonant tongue movements to a greater degree in fast speech (Engstrand, 1988). In fast speech, diphthongized vowels become more central overall, and the target qualities may not be fully reached (Kuronen, 2000, p. 37). Boundary tones are also detrimental in this study since word level pitch accents are studied, and these are generally overwritten by IP-tones. On the other hand, the sentences in the list were designed so that the target words are not in initial or final positions; the target words in the short story are found in the beginning of an intermediate phrase (ip) 5 times and at the and of an ip 11 times.

3.3 Participants

14 adult speakers of Central Swedish were recruited through informal social networks. The participants were required to be raised in Stockholm county and speak Swedish as an L1. No recruited participant deviated dialectically in their production of vowels or word accents. All participants signed informed consent forms, as detailed in section 3.4. Among the participants, the distribution of centralized /iː/ and /yː/ (both described as [ɨː] in Persson (2025)) was overwhelming as only three participants produced the more front articulations [iː] and [yː].

3.4 Ethical considerations

The ethical considerations relevant for this study are two-fold: participants' consent and treatment of sensitive personal data. The participants were informed about the procedure of the recording beforehand, and all participants signed a consent form confirming that they had taken part in the following information about the study: the aim of the study, investigating vowel acoustics and word accents in their dialect; the procedure of the recording, as detailed above; that participation was completely voluntary, and that they could cancel their participation at any time; and that all data would be anonymized.

As it pertains to the treatment of sensitive personal data, the data collected for this study has been deemed to not be sensitive. According to the Swedish Authority for Privacy Protection (Integritetsskyddsmyndigheten [IMY], 2026), sensitive personal data concerns data of a persons racial or ethnic origin, political opinions, religious or philosophical beliefs, membership of a trade union, health, a person's sex life or sexual orientation, genetic data, and biometric data that is being used to uniquely identify a person. No data on the identity or affiliation of the participants have been collected, apart from the prerequisite that they speak Central Swedish. This dialect is not directly connected to racial or ethnic origin. No genetic data has been collected for this study. The data collected for this study consists of the voice recording of the participants. No technology or other tools for voice recognition or any other way to associate the recorded voices to the participants have been used in this study. Thus, it has been deemed that the present data is not of a sensitive personal quality.

3.5 Pilot study

A pilot study was conducted to select a suitable elicitation method. The choice was between a reading task with a list of sentences, and a reading task with a short story. The material for both can be found in the appendix. Two L1 speakers of Central Swedish participated as part of the pilot. During the recording, they were first asked to read the list of sentences, then the short story. Otherwise, the procedure was identical to the one used throughout the study, as detailed in section 3.1. The participants of the pilot study spoke for around four and a half minutes each, including rereading certain sentences. The reading of the short story in the pilot was incorporated in the study. This data was deemed usable since it was comparable to the data from the other participants.

The pilot showed that the more useful method for this study was the reading of the short story. This is for two reasons: (1) because the participants tended to read the list of sentences faster; (2) because the format of a list of sentences triggered intonational phrase (IP) boundary tones, i.e. H%, which were more disruptive than the ip tones of the short story.

The differences between the two microphones used in the pilot study were marginal, but the DPA 4066 omnidirectional headset mic was preferred because there was less noise in general in those recordings. The recordings from this microphone are what the acoustic analysis is based on.

3.6 Acoustic analysis

The acoustic analysis was carried out in Praat (Boersma & Weenink, 2024). Segment boundaries for the target vowels were annotated to TextGrid manually by the author, an L1 speaker of Central Swedish. Following Persson (2025) and Pelzer & Boersma (2019), the formants were calculated with the Burg algorithm in praat. The algorithm was paramaterized with a time step of 0.01 s, a window length of 0.025 s, and with pre-emphasis applied from 50 Hz. The settings of maximum number of formants and formant ceiling used for the algorithm depended on the speaker. Usually, the conventional settings (5 formants; 5500 Hz and 5000 Hz for women and men respectively) worked well, but certain speakers' formants were more optimally calculated with the maximum number of formants set to 4.5 or 5.5[1], and the speech of two female speakers were more reliably calculated with the formant ceiling set to 5000 Hz. For certain vowel qualities, namely [uː] and (to a lesser extent) [aː], the formant settings had to be adjusted further for multiple speakers.

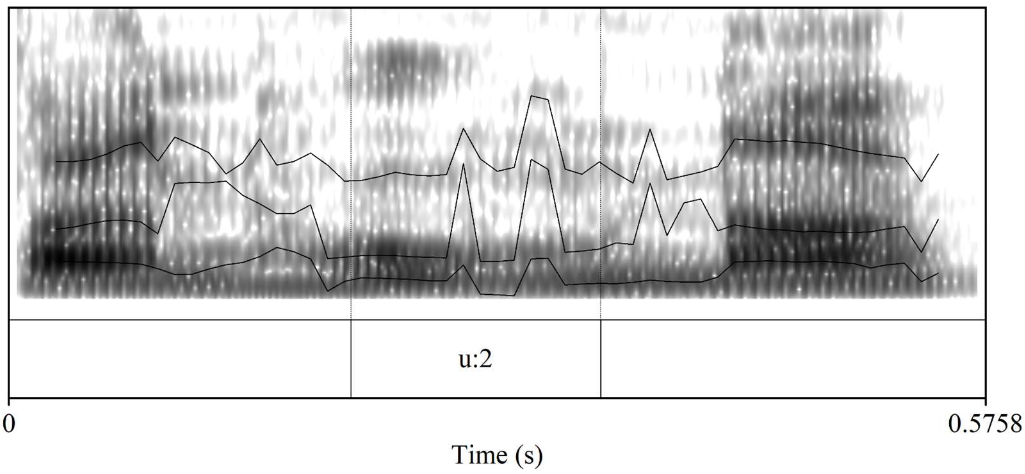

Figure 1: Spectrogram, annotation, and formant tracks with maximum formant number set to 5. Formant calculation does not correspond to actual formant values: F1 was not found at multiple time points.

[1] The algorithm does not search for "half" formants. The setting "Maximum number of formants" can be set to multiples of 0.5 because the Burg algorithm in praat uses the "Maximum numbet of formants" variable multiplied by 2 when calculating (Boersma & Weenink, 2024).

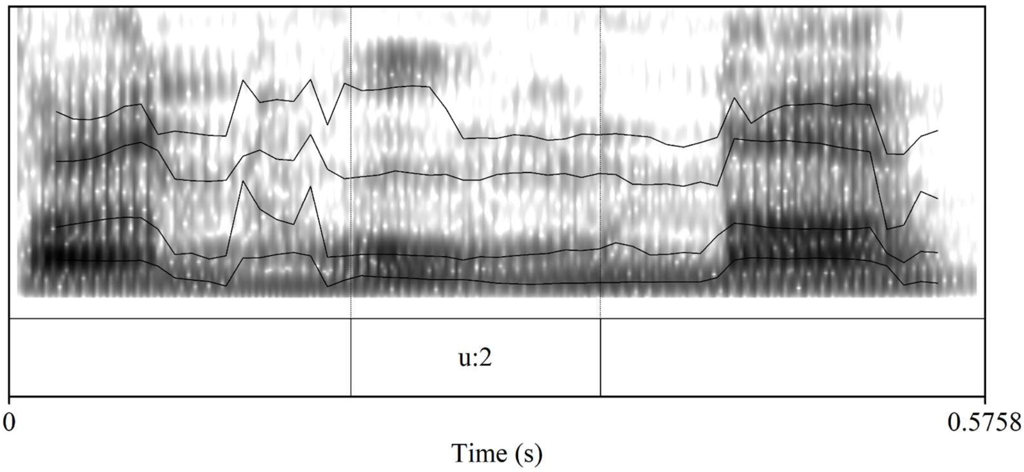

Figure 2: Spectrogram, annotation, and formant tracks with formant number set to 5.5. Formant calculation now corresponds with actual formant values in the target vowel.

Formants were calculated at five time points: 20 %, 35 %, 50 %, 65 %, and 80 % into the vowel. Formant values of the first three formants were extracted using a praat script written by Matthew Winn (Listen Lab, 2021), to then be compiled.

The acoustic analysis of the interaction between F0 and formant movements was mainly done by eye. Pitch tracks were found using the filtered autocorrelation analysis method in praat (Boersma & Weenink, 2024), with the pitch range settings adjusted for different speakers and words. Pitch range was adjusted until a pitch track with no octave jumps or other unexpected fluctuations was found.

4. Results and analysis:

Diphthongization was present in every long vowel, to different degrees. The noticeably diphthongized vowels in the material were /eː/ and /oː/. The long vowels /uː/, /ʉː/, and /oː/ often exhibited the insertion of an approximant within the duration of the vowel, usually near the end of the vowel. Though approximant insertion was slightly rarer in /oː/, one can be seen in figure 7.

The short vowels displayed diphthongization to a much lesser degree than the long vowels.

4.1 Formant dynamics and F0: accent I & II

In general, the dynamics of F0 on the target vowels were consistent across speakers. Figures 3 through 9 illustrate the general patterns of F0 and formant movements of diphthongized vowels found in the material. The most common tonal pattern for accent I words is the one exemplified in figure 3, and the most common tonal pattern for accent II words is the one exemplified in figure 6.

Though the analysis focuses on the target vowels, figures 3-9 also include the whole word containing the target vowel.

In the material, if the second H tone of an accent II tonal contour is not found during the diphthongized vowel, it is written within parentheses, i.e. H*L(H). This maintains the distinction of focus, as accent II without focus is written as H*L.

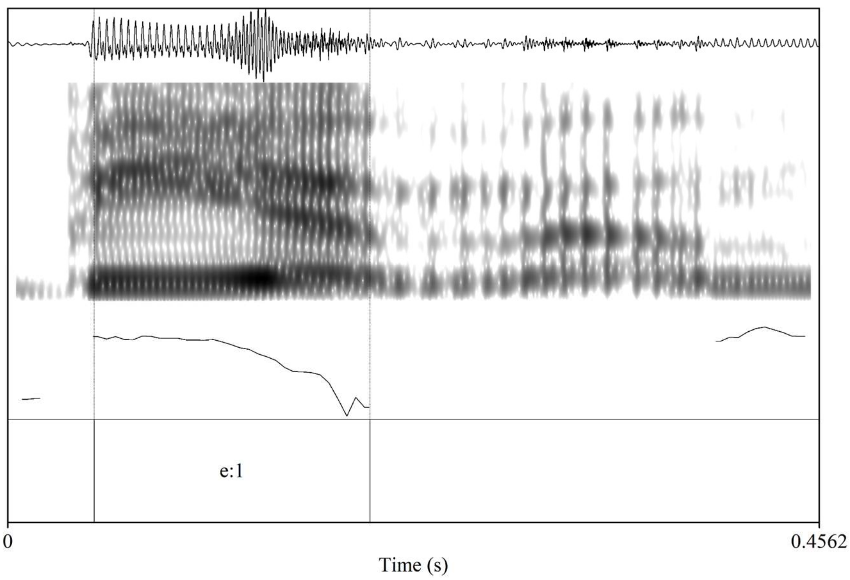

Figure 3: Soundwave, spectrogram, pitch track, and annotation of 'benen', the target word for /e:/ in accent I. Tonal pattern of focus surfaces as HL. Vocalic steady-state component followed by an opening offglide with a distinct onset.

Focused accent I words often displayed a somewhat falling tonal contour rather than the expected rising tonal contour (see table 2) When F0 displayed this HL "plateau-fall" pattern, the beginning of the fall usually coincided with the onset of a diphthongized offglide component. Figure 3 illustrates a typical F0-pattern found in the data. In diphthongized vowels with this tonal pattern, both F0 and the formants begin to move approximately simultaneously, around the middle of the vowel. Figures 3 and 4 also show that this pattern appeared in both word accents, as long as the F0-contour was similar within the diphthongized vowel.

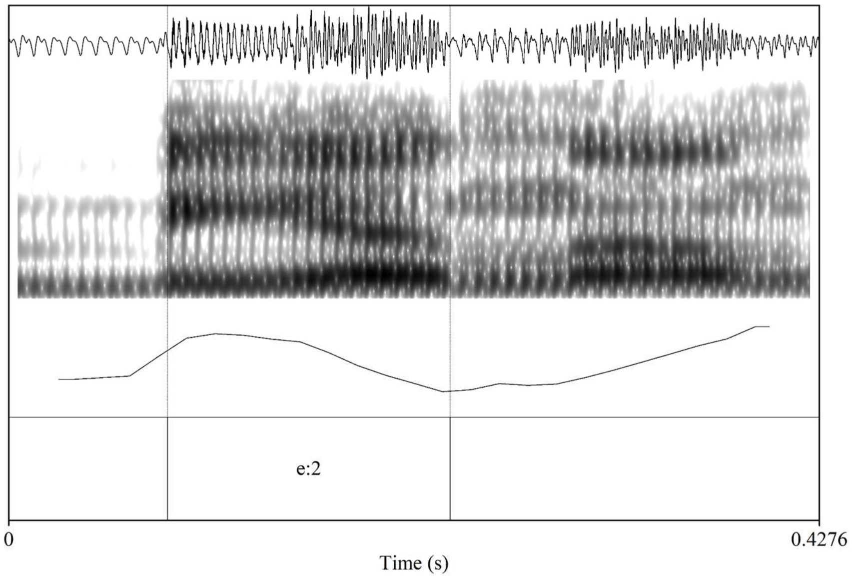

Figure 4: Soundwave, spectrogram, pitch track, and annotation of ‘benan’, the target word for /e:/ in accent II. Tonal pattern of focus surfaces as H*L(H). Vocalic steady-state followed by an opening offglide with a distinct onset.

Figure 4 exemplifies a diphthongized vowel in an accent II word with sentence focus. A similar tonal pattern as the one found in figure 3 surfaces: a relatively unchanging plateau until around the middle of the vowel, where F0 begins to fall. The expected word accent tonal contour surfaced as H*L within the target vowel and H on the second syllable.

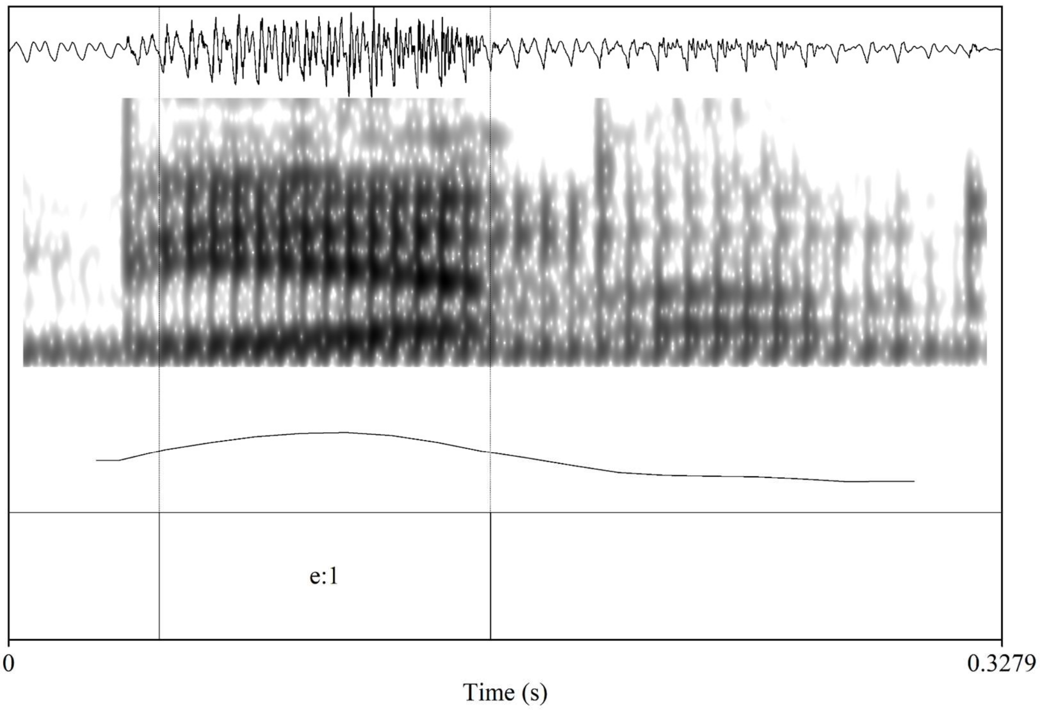

Figure 5: Soundwave, spectrogram, pitch track, and annotation of ‘benen’, the target word for /e:/ in accent I. Tonal pattern of focus surfaces as L*H. Vocalic steady-state, gradually opening with no distinct onset.

Focused accent I words displayed the expected L*H tonal contour, though more rarely than the falling contour. Figure 5 illustrates the rising tonal contour within the diphthongized vowel. The formant dynamics of the diphthongized vowel are more gradual, with no distinct boundary between steady-state and offglide. Increase in formant movements might coincide with the F0 turning point, specifically for F1.

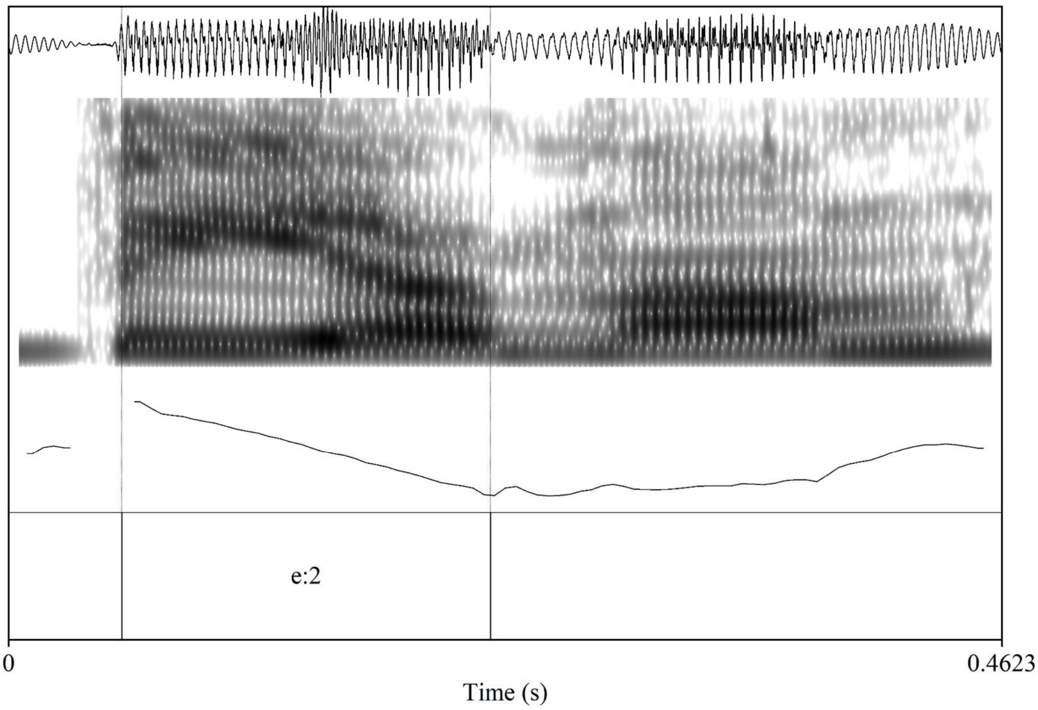

Figure 6: Soundwave, spectrogram, pitch track, and annotation of ‘benan’, the target word for /e:/ in accent II. Tonal pattern without focus surfacing as H*L. Vocalic steady-state followed by an opening offglide with a distinct onset.

When accent II words were produced without focus, they typically displayed the tonal pattern illustrated in figure 6. The distinct boundary between steady-state and glide in the diphthongization of /e:/ shows no correlation with the F0 contour, whose fall was steady throughout the vowel.

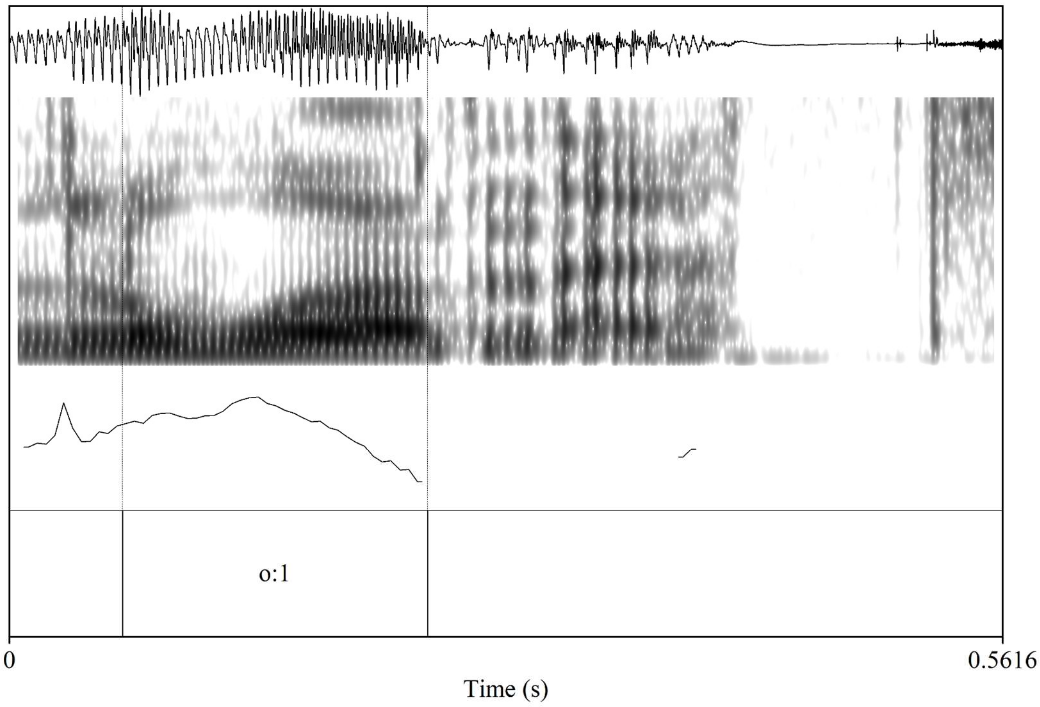

Figure 7: Soundwave, spectrogram, pitch track, and annotation of 'lånet', the target word for /o:/ in accent I. Tonal pattern of focus surfaces as L*H. Vocalic steady-state followed by an opening offglide. The steady-state is interrupted by an approximant.

Figure 7 illustrates the expected F0 contour of accent I in focus. Of note, however, is the dip in F0-curve. This dip occurs at the same time there is visible dampening in the sound-wave. This dampening coincides with the audible insertion of a labial approximant. Insertion of labial approximant was also widespread for /uː/ and /ʉː/.

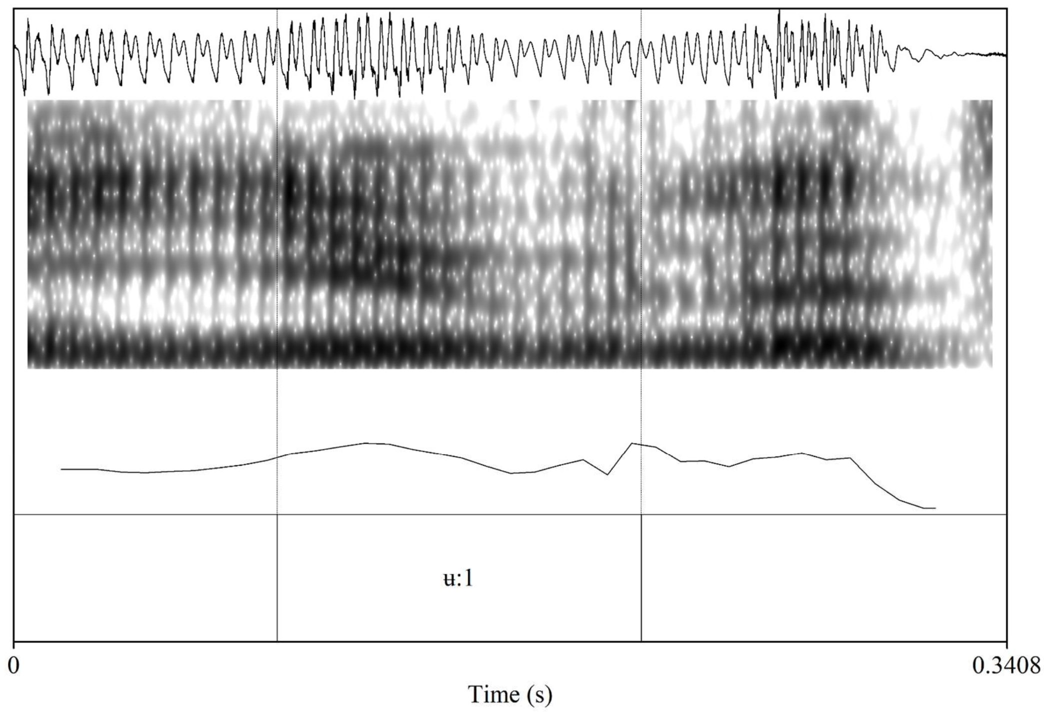

Figure 8: Soundwave, spectrogram, pitch track, and annotation of 'hjulet', the target word for /ʉ:/ in accent I. Tonal pattern of focus surfaces as a flattened L*H. Diphthongized vowel with a characteristic insertion of an approximant.

The insertion of a labial approximant mentioned above usually followed the pattern illustrated in figure 8: Approaching the end of the vowel, amplitude and frequency of the speech wave was dampened, followed by a release phase before the end of the vowel. The pitch track shows that F0 falls slightly, even before the approximant, and then continues falling during the insertion. During the release phase, F0 rises back to the level it was at before falling.

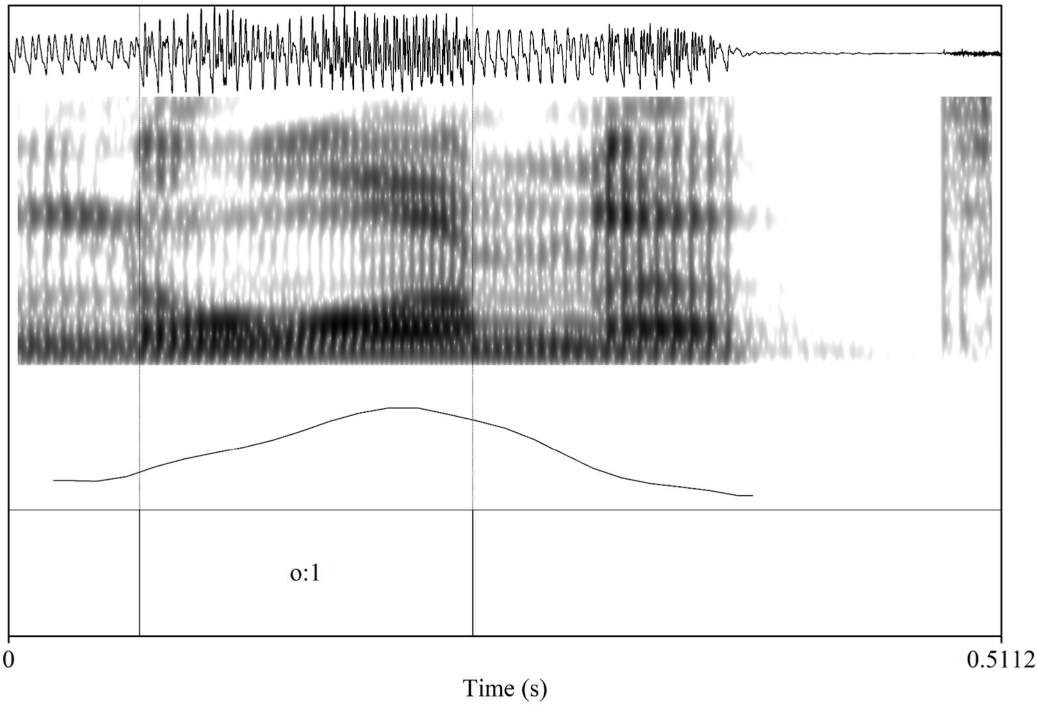

Figure 9: Soundwave, spectrogram, pitch track, and annotation of ‘lånet’, the target word for /o:/ in accent I, produced by speaker 6. Tonal pattern of focus surfaces as L*H. Vocalic steady-state followed by a short opening offglide.

Figure 9 shows how F0 and formant movements may be intertwined. During the rising F0, the formants move minimally. As F0 reaches its maximum during the vowel, and especially as it starts to fall around 80 % into the vowel, F1 and F2 move drastically. This is most apparent when looking at the extracted formant values, as shown in table 4 below.

	20 %	35 %	50 %	65 %	80 %
F3	2663	2690	2773	2791	2769
F2	750	725	696	763	947
F1	384	364	383	416	580

Table 4: Formant values of /o:/ in accent I, produced by speaker 6.

4.2 Formant dynamics and creaky voice: accent I & II

Three speakers heavily featured creaky voice phonation throughout the recordings: speakers 9, 13, and 14. Speaker 9 featured the most creaky voice, followed by speaker 13 and

14, respectively. The target vowels that featured creaky voice to the greatest extent were /eː/, /oː/, /uː/, /o/, and /ʉ/. Spectrograms of these vowels,produced by these speakers, are illustrated in figures 10 to 14 below. The data showed that creaky voice tends to correspond to accent I to a greater extent than accent II.

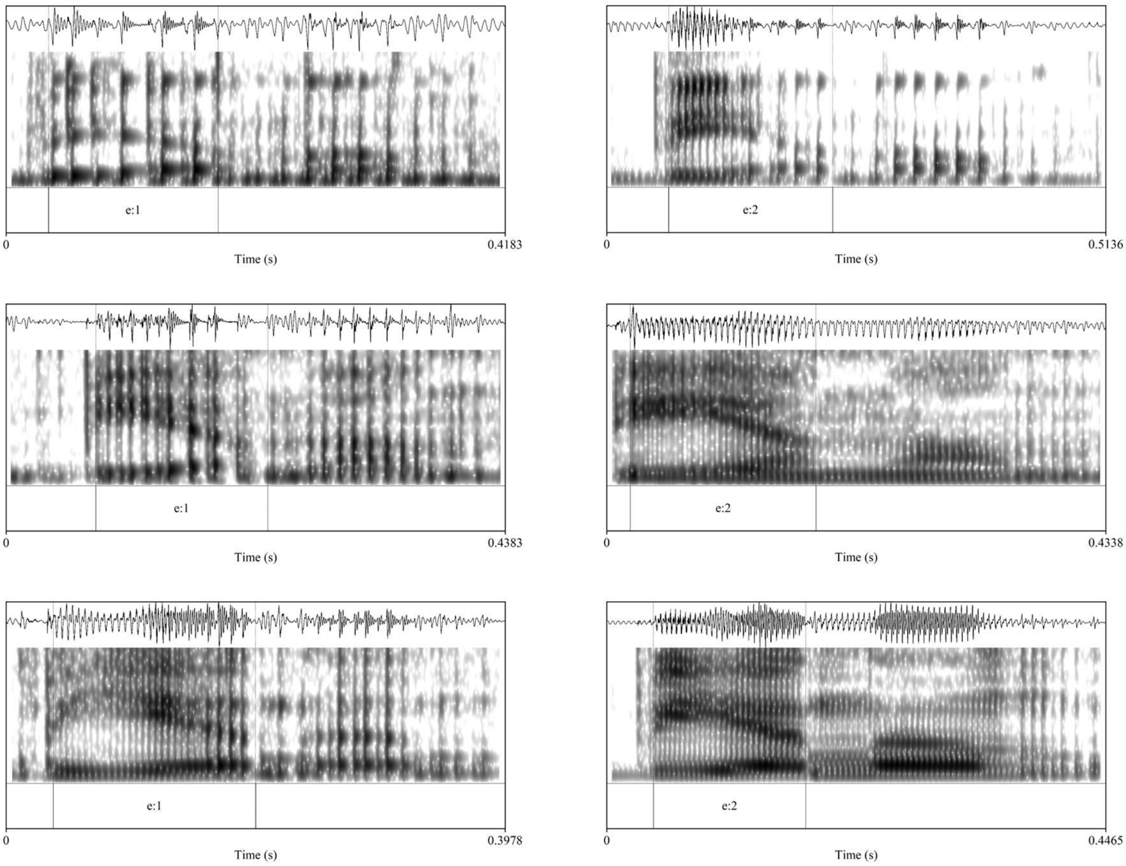

Figure 10: Soundwaves, spectrograms, and annotations of /eː/ in accent I (left) and accent II (right) of speakers 9 (top), 13 (middle), and 14 (bottom).

For /eː/, the accent I words featured creaky voice during the diphthongized vowel for all three speakers. In accent II, creaky voice appeared further into the word. For speaker 9, creaky voice was shifted to the offglide component of the diphthongized vowel, while for speakers 13 and 14, creaky voice shifted to the coda of the second syllable. Both target words for /eː/ appeared in the middle of a sentence but at the end of a main clause (see appendix).

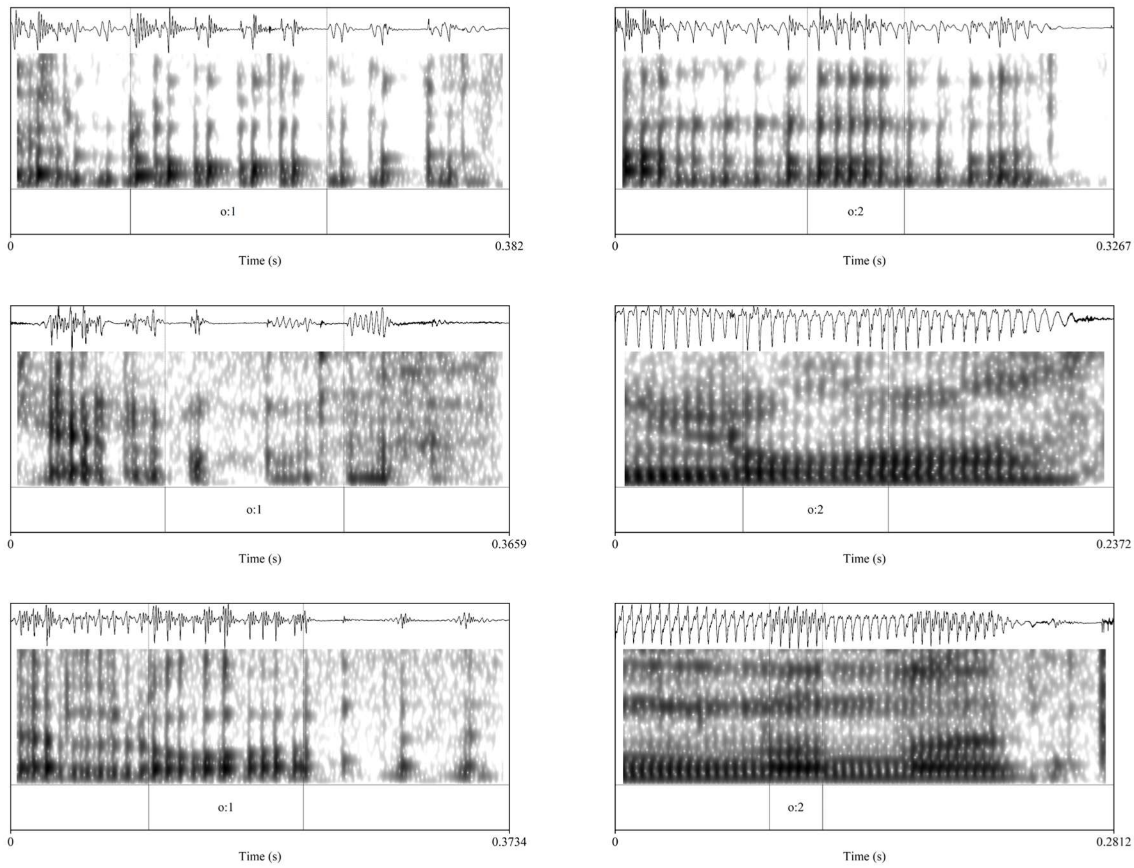

Figure 11: Soundwaves, spectrograms, and annotations of /o:/ in accent I (left) and accent II (right) of speakers 9 (top), 13 (middle), and 14 (bottom).

For /o:/, the accent I words featured creaky voice throughout the whole word. The diphthongized vowel was highly aperiodic for speaker 13. For speakers 13 and 14, the accent II words featured no creaky voice during the target vowel or anywhere else in the word. Speaker 9 featured creaky voice in the accent II word. However, there was greater periodicity in the target vowel compared to both the phonation in the accent I word and the phonation in the accent II word outside of the target vowel. The target word for accent I appeared at the end of a sentence, while the target word for accent II appeared in the middle of a clause (see appendix).

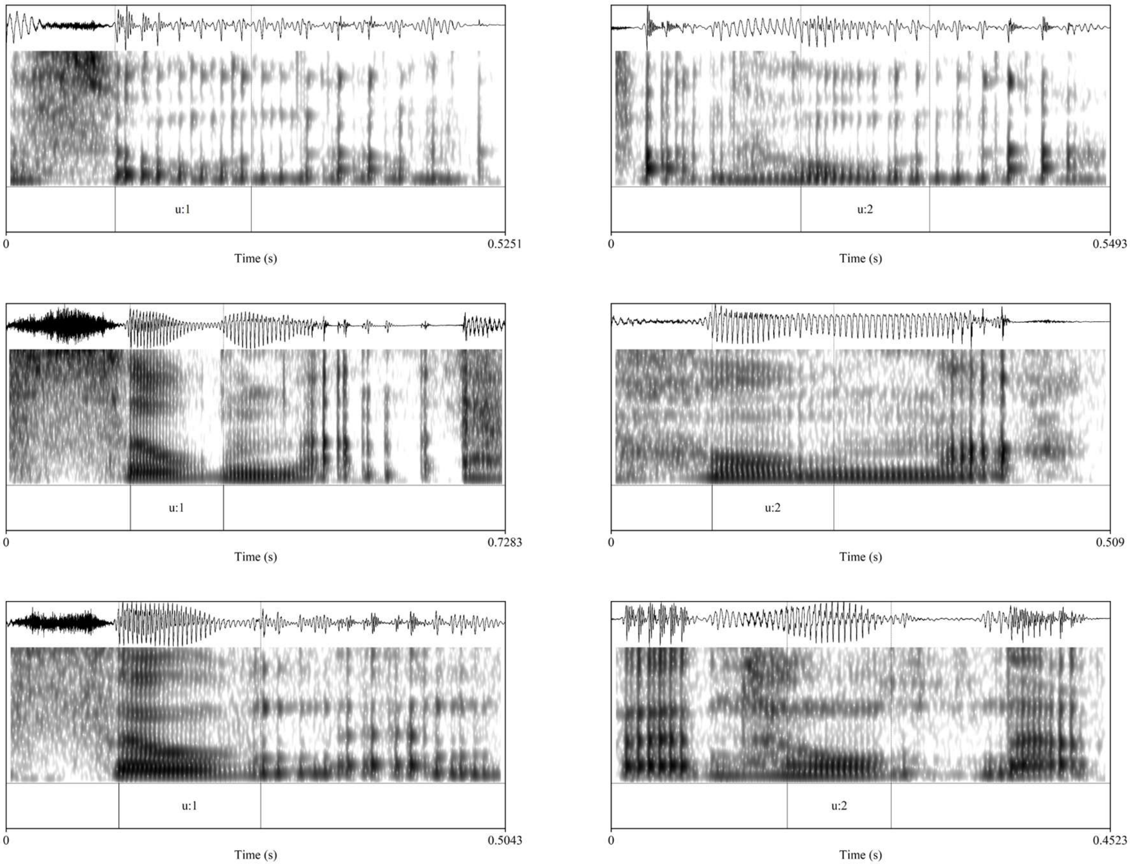

Figure 12: Soundwaves, spectrograms, and annotations of /u:/ in accent I (left) and accent II (right) of speakers 9 (top), 13 (middle), and 14 (bottom).

For /u:/, the accent I words featured creaky voice throughout the diphthongized vowel as well as the following segments for speaker 9, and immediately after the target vowel for speaker 14. For speaker 13, creaky voice appeared only during the second syllable for both accent I and accent II. The creaky voice of speaker 14 had less periodicity in the accent I word than the accent II word. For speaker 9, creaky voice in the accent II word appeared half-way through the diphthongized vowel. For speaker 14, creaky voice in the accent II word appeared at the end of the vowel in the second syllable. The target word for accent I appeared in quotation marks in the middle of a sentence, while the target word for accent II appeared at the end of a sentence.

Some part of the target vowel in the accent II word for speaker 9 may have reduced creaky voice. Creaky voice appeared before the target word, and reappeared during the latter part of the diphthongized vowel in the target word. Between these periods of creaky voice lies the stressed vowel of an accent II word, which features modal voice for the most part.

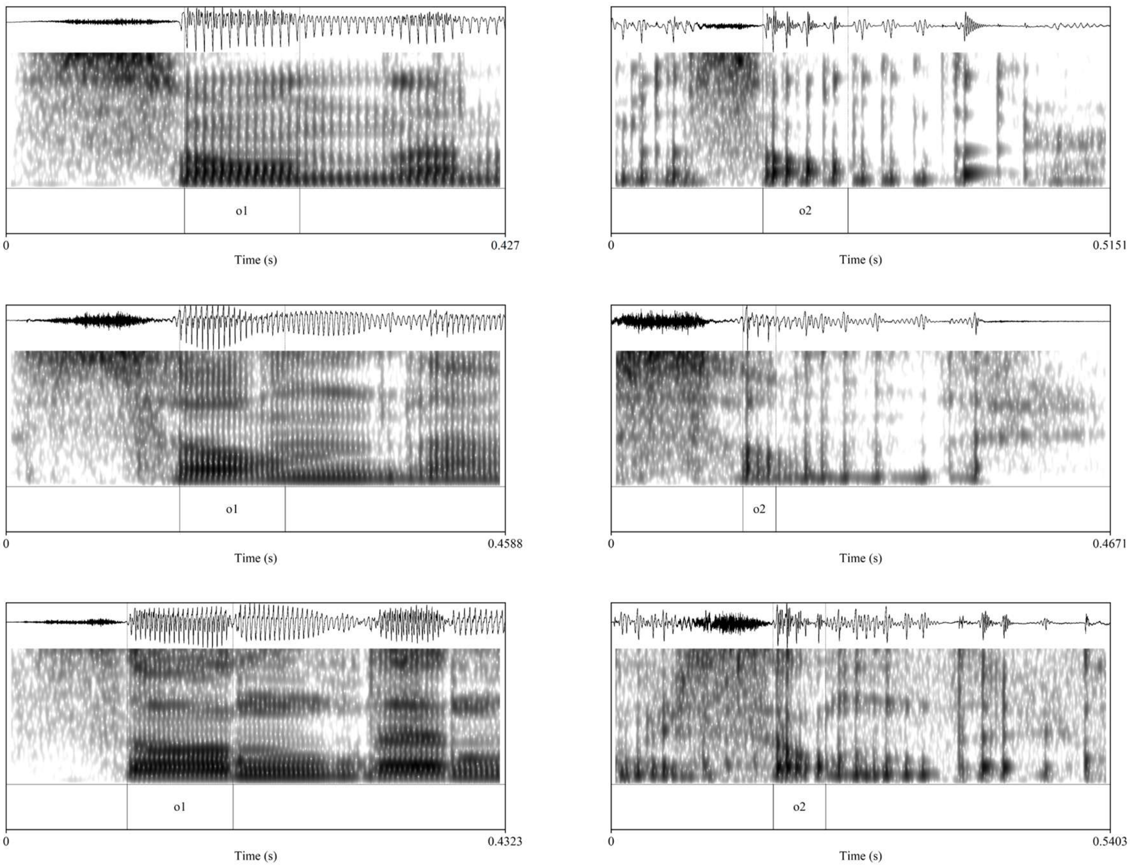

Figure 13: Soundwaves, spectrograms, and annotations of /o/ in accent I (left) and accent II (right) of speakers 9 (top), 13 (middle), and 14 (bottom).

The short vowels /o/ (figure 13) and /ʉ/ (figure 14) break the pattern seen in the long vowels above: The accent II words featured more creaky voice than the accent I words. For all three speakers, the target vowels in the accent I words feature modal voice. The same goes for the second syllable. All three speakers feature creaky voice in the accent II words. Comparing the phonation in the target vowel to second syllable and the coda of the first syllable, there is more periodicity during the vowel in the first, stressed syllable.

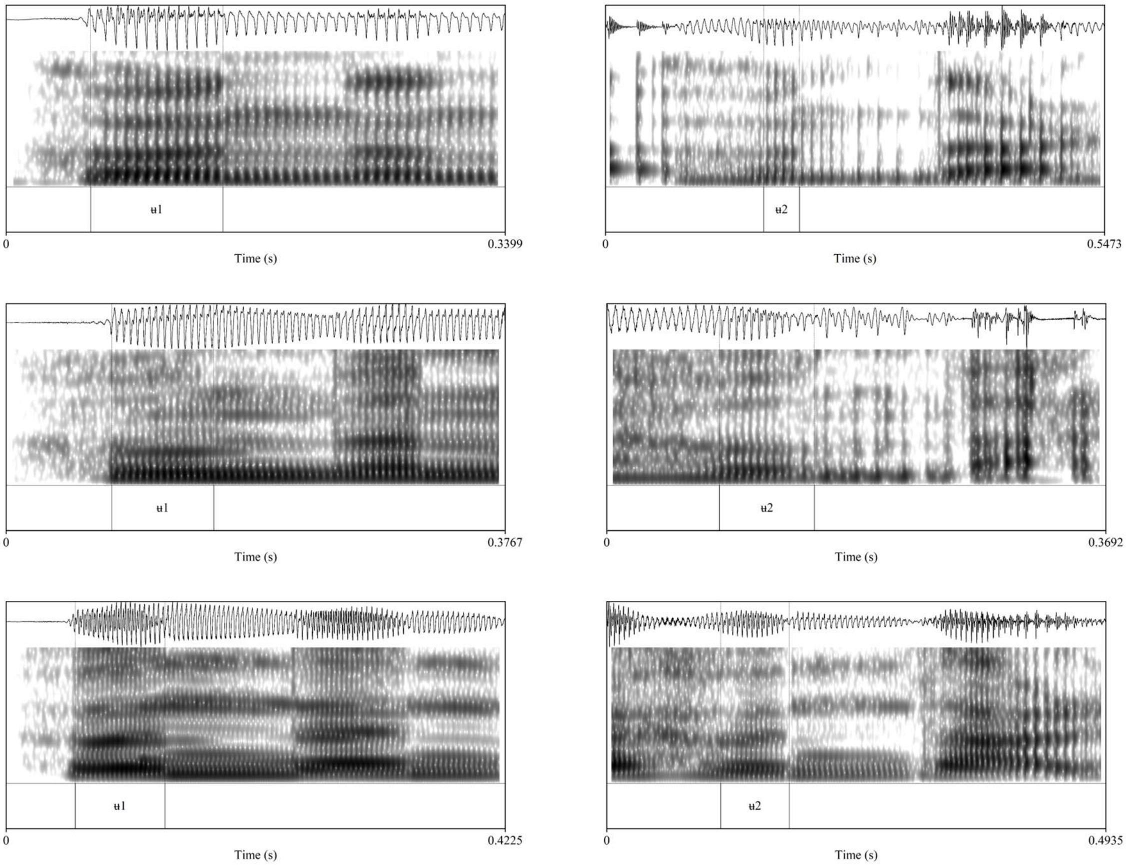

Figure 14: Soundwaves, spectrograms, and annotations of /ʉ/ in accent I (left) and accent II (right) of speakers 9 (top), 13 (middle), and 14 (bottom).

Also in the target words for /ʉ/ did accent II feature more creaky voice than accent I. The accent I words in figure 14 featured only modal voice for all three speakers. In the accent II words, creaky voice appeared in the consonant coda of either the stressed syllable (speakers 9 and 13) or the second, unstressed syllable (speaker 14). For speaker 9, creaky voice in accent II falls into the same pattern as it did for speaker 9 with the /uː/ vowel. Creaky voice appears before the stressed syllable, disappears, and then reappears during the stressed syllable. The difference between /uː/ and /ʉ/ is that creaky voice reappears during the consonant coda for the short vowel.

4.3 Formant dynamics in accent I vs accent II

The following figures display a selection of vowel formant data that illustrate differences in vowel dynamics between the two word accents. The figures show values of the first three formants at five time points into the vowels. The values are averages across all speakers and the error bars display the standard deviation.

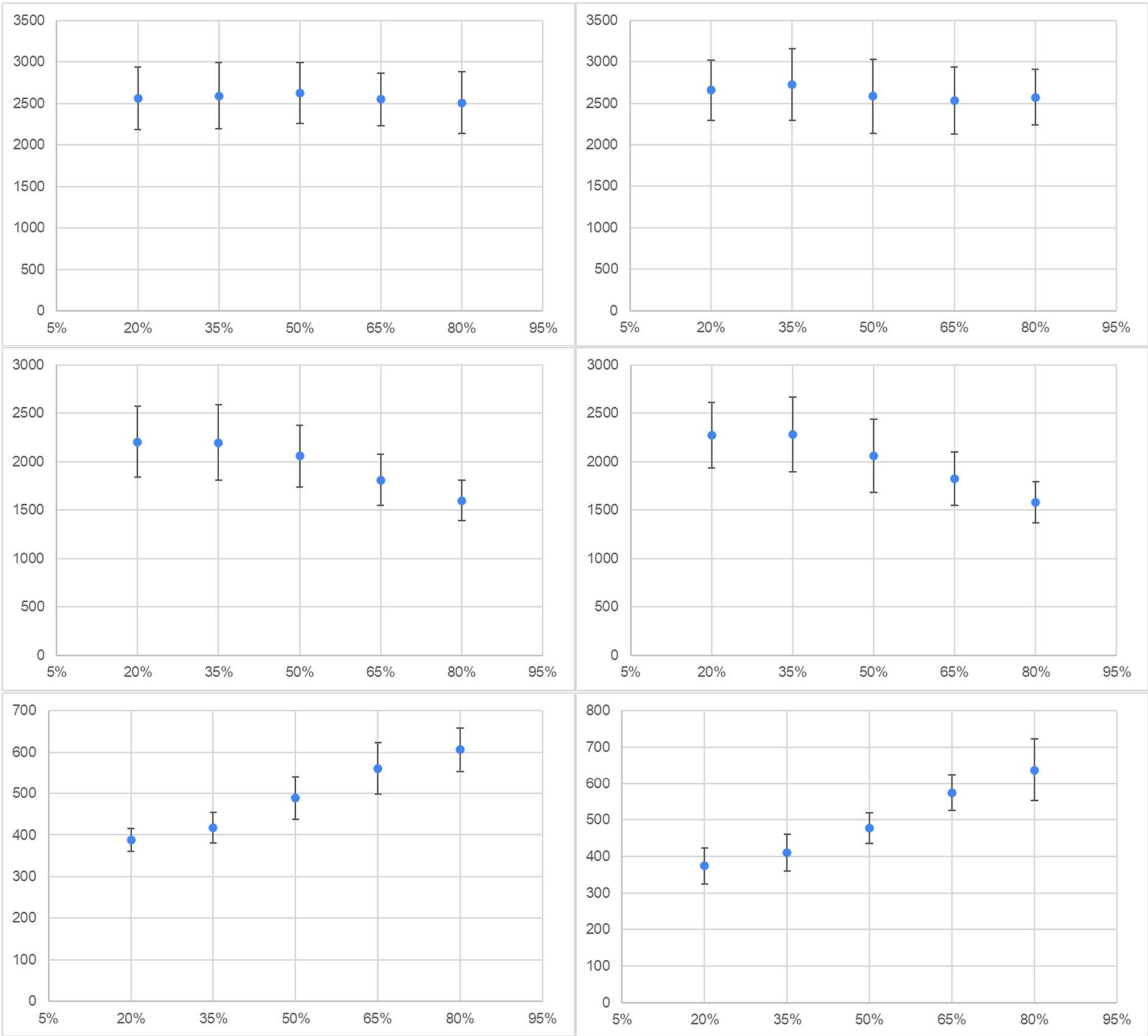

Figure 15: Formant values at five time points into /e:/,
divided by word accent (accent I to the left, accent II to the right) and formant (F3 at the top, F2 in the middle, F1 at the bottom).

Figure 15 shows that the vowel dynamics of /e:/ are almost identical between accent I and accent II, especially for F2. Nonetheless, F1 differs slightly between the two word accents: In accent II, the average F1 rises more between time points 50 % and 65 %, and between time points 65 % and 80 %. Furthermore, at 80 % into the vowel, the standard deviation is greater.

Figure 16: Formant values at five time points into /ʉː/, divided by word accent (accent I to the left, accent II to the right) and formant (F3 at the top, F2 in the middle, F1 at the bottom).

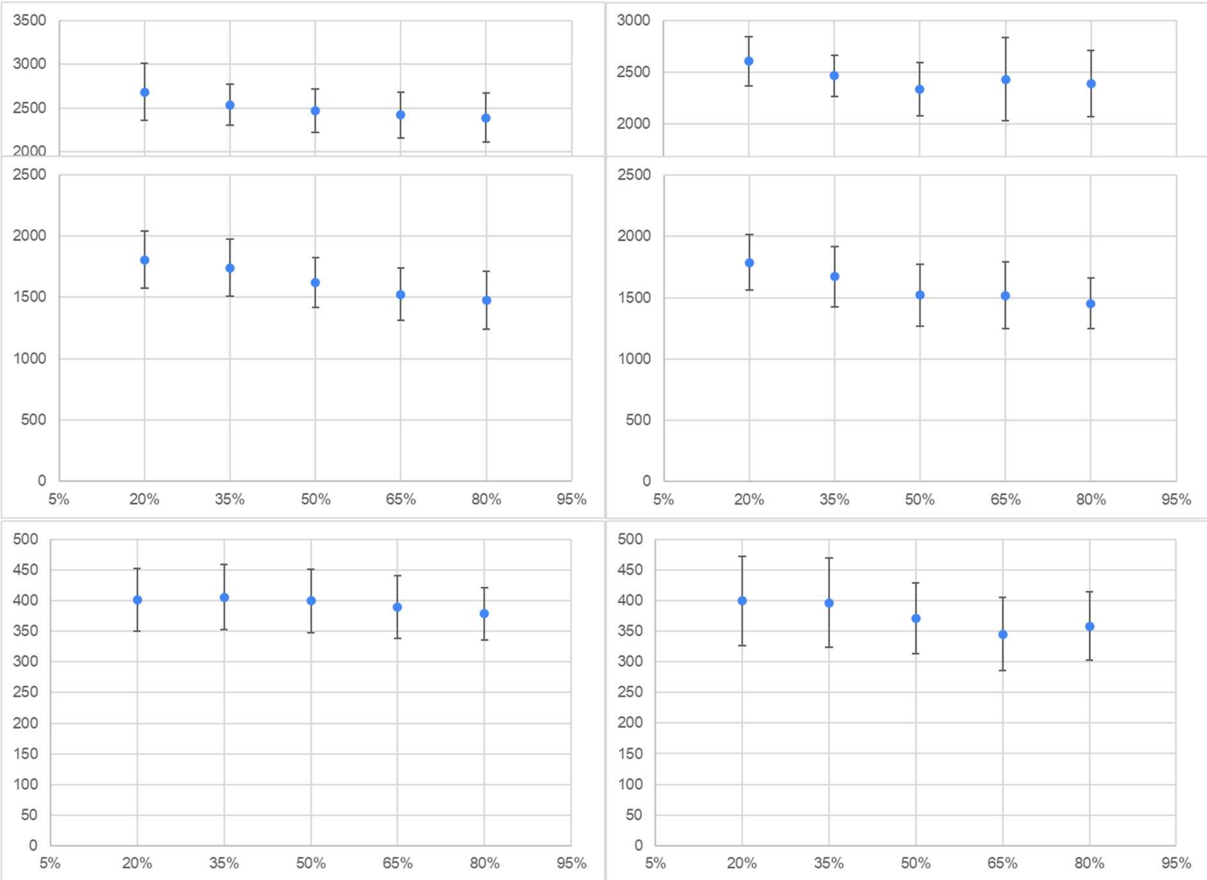

Figure 16 shows that the dynamics of F1, and F2 to a lesser extent differs between word accents for /ʉː/. The average F1 stays around 400 Hz in accent I, falling slightly near the end, and average F2 falls gradually for the full duration of the vowel.

In accent II, the average F1 falls between time points 35 % and 65 %, and rises between time points 65 % and 80 %. Average F2 falls between time points 20 % and 50 %, and then stabilizes.

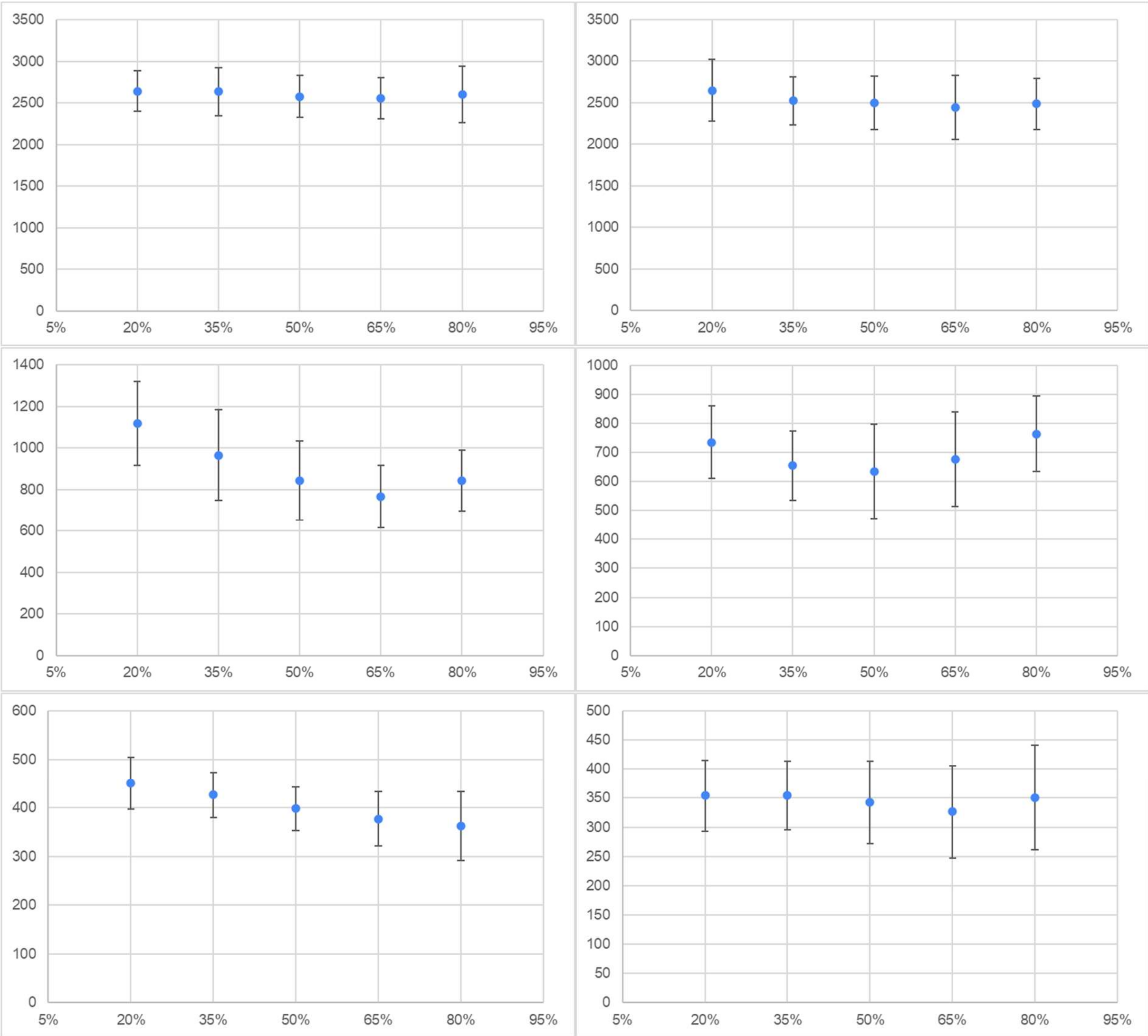

Figure 17: Formant values at five time points into /u:/, divided by word accent (accent I to the left, accent II to the right) and formant (F3 at the top, F2 in the middle, F1 at the bottom).

Figure 17 shows the big differences between /u:/ vowel dynamics and formant values between the two word accents. In accent I, the F1 of /u:/ falls steadily from an average of 450 Hz to an average of 362 Hz. Between 20 % and 65 % into the vowel, average F2 falls from around 1100 Hz to 770 Hz. Between time points 65 % and 80 % into the vowel, F2 rises back to where it was at 50 % into the vowel.

In accent II, the average F1 falls around 50 Hz between the time points 50 % and 65 %, and rises between the time points 65 % and 80 % into the vowel. Average F2 falls during the first half of the vowel, and rises during the second half. F2 falls further in accent I than in accent II.

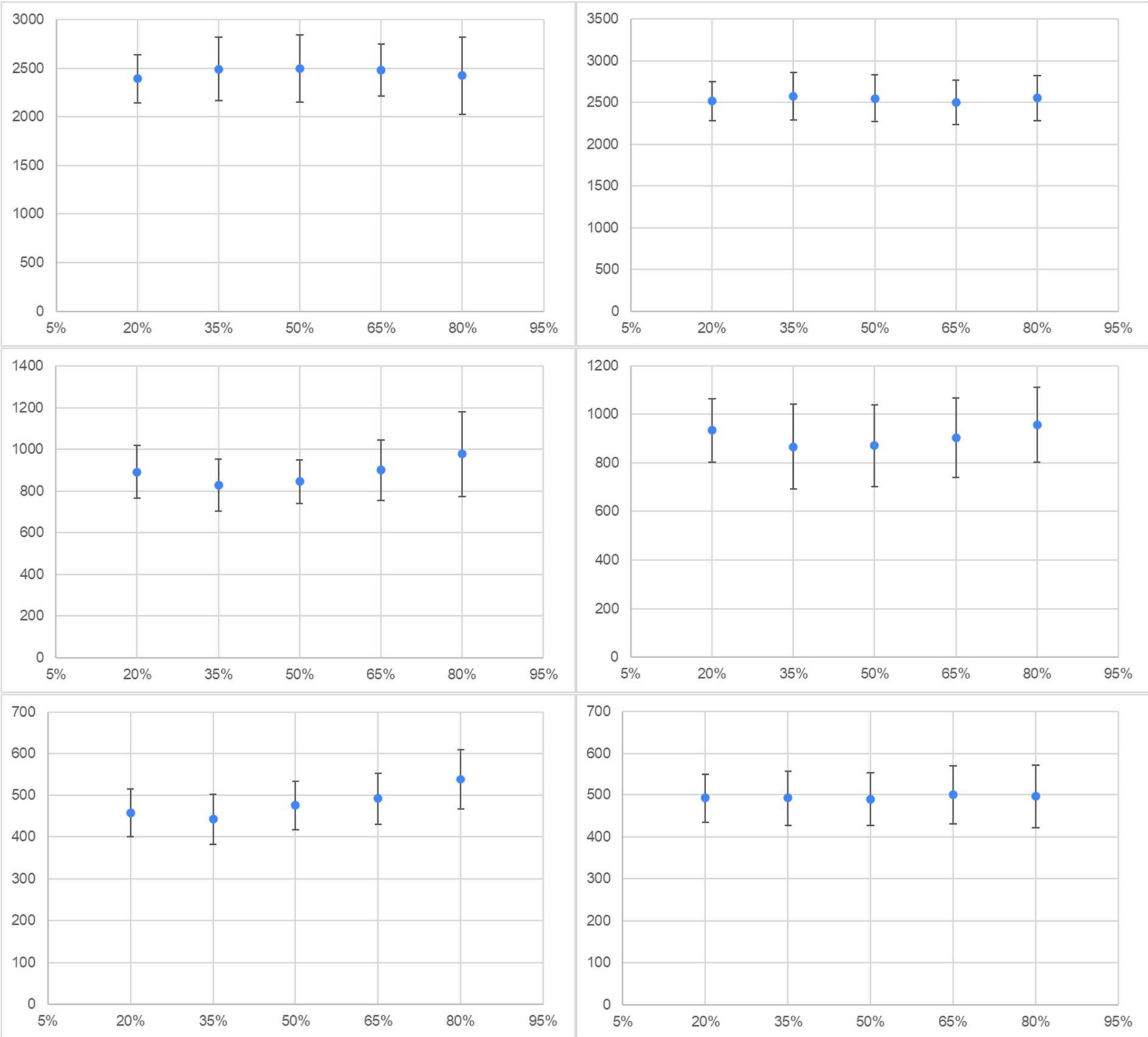

Figure 18: Formant values at five time points into /oː/, divided by word accent (accent I to the left, accent II to the right) and formant (F3 at the top, F2 in the middle, F1 at the bottom).

Figure 18 illustrates the dynamic spectral differences between the two word accents in /oː/. In accent I, F1 rises between time points 35 % and 50 %, and 65 % and 80 %; and falls between time points 20 % and 35 %. F2 follows a similar pattern to F1, though the rise between time points 35 % and 50 % is smaller.

In accent II, F1 stays relatively unmoving just below 500 Hz. F2 falls between time points 20 % and 35 %, and rises between time points 50 % and 65 %, and between time points 65 % and 80 %. The differences in formant values between the time points of F2 are smaller in accent II compared to the ones in accent I.

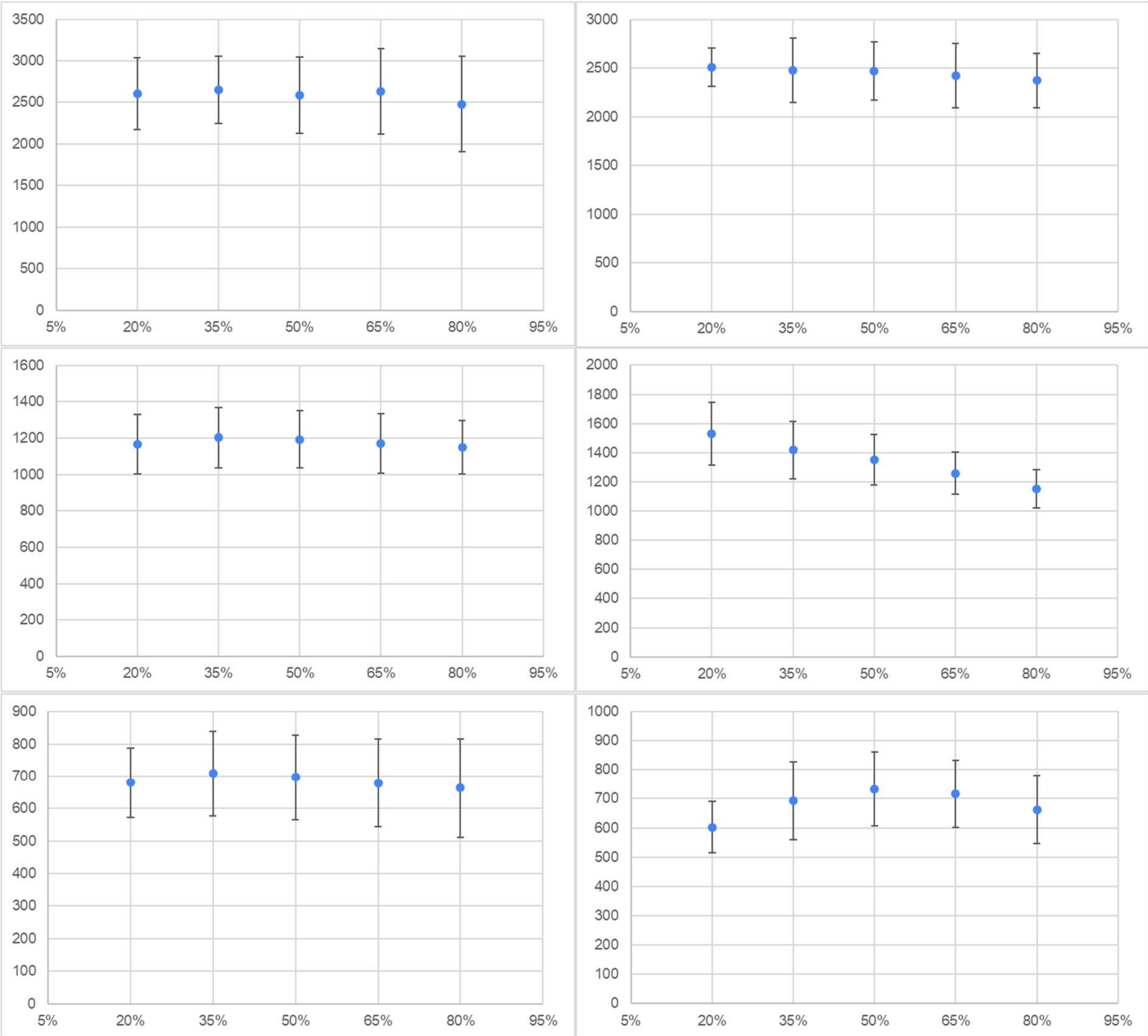

Figure 19: Formant values at five time points into /a/, divided by word accent (accent I to the left, accent II to the right) and formant (F3 at the top, F2 in the middle, F1 at the bottom).

Figure 19 illustrates the dynamic spectral differences between the two word accents in short /a/. For accent I, F1 rises between time points 20 % and 35%, and falls steadily between time points 35 % and 80 %. F2 follows the same pattern, but less drastically, staying somewhat close to 1200 Hz.

For accent II, F1 rises sharply between time points 20 % and 35 %, rises further between time points 35 % and 50 %, and falls between time points 50 % and 80 %. The average F2 falls steadily from 1530 Hz at 20 % into the vowel to 1160 Hz at 80 % into the vowel.

6. Discussion:

In light of the research questions (section 1.1), several points can be made.

When movements of F0 are not aligned with diphthongization at all—e.g. in the H*L tone of accent II without focus (figure 6)—F0 and formants do not interact. This is to be expected. In the cases when they are aligned, the results show that they indeed interact. With a rising pitch contour, opening diphthongized vowels were produced more monophthongized (figure 5), which aligns with the findings of Li & Al-Tamimi (2024). The labial approximant insertion found in closing diphthongized vowels with rising pitch contours could be argued to be an additional closing gesture. In that case, it might coincide with rising pitch contours (as seen in figure 7), though the dampening of the approximant may affect the acoustic realization of rising F0 (figure 8). The comparison between accent I and accent II (figure 15) show that the falling tonal gesture of accent II might facilitate opening diphthongization to a greater degree than accent I. Thus, the results arguably indicate that the dynamics of F0 affect the dynamics of vowel formants, where rising F0 facilitates closing diphthongization and falling F0 facilitates opening diphthongization.

With a falling pitch, the onset of opening offglides consequently coincided with the onset of falling F0 (figures 3 and 4), as well as the turning point where F0 goes from rising to falling (figures 5, 7, and 9). Thus, the results indicate that the movements of a dynamic F0 tend to coincide with formant movements in diphthongized vowels.

Because F0 and formants interact, and certain ranges of F0 coincide with creaky voice, the above indications imply that creaky voice may exhibit similar characteristics. However, the results of this study do not necessarily indicate clear interactions between formant movements and creaky voice. This is partly because of the limited data set of only three speakers; very few items had the onset of creaky voice within a target vowel. The data with creaky voice starting during the vowel segment (figure 10: top right, middle left, and bottom left spectrograms; figure 12: top right spectrogram) do not paint a coherent picture either way. Thus, the question whether creaky voice coincides with formant movements still needs further investigation.

Comparing accent I and accent II more holistically, the first thing to note is that F3 displayed very minimal differences between the two word accents. Secondly, spectral differences between adjacent time points tended to be greater for vowels in accent II. In addition, the values of accent II generally had greater standard deviation, as can be seen in figures

15, 16, 17, and 18. This implies that accent II might facilitate more diphthongization, but that this facilitation might not apply to the same degree for all speakers.

Exception to the above are found in /oː/ and /uː/, where F1 in accent II is more monophthongized, which weakens the argument for the implication above. There might be some way to explain why the non-high back vowels specifically would be exempt from the facilitation of diphthongization in accent II, but that would be a task for future phonological research.

Thus, I would say that accent II in Central Swedish facilitates vowel dynamics and diphthongization in the first vowel of two-syllable words with stress on the first syllable, more than accent I.

Furthermore, the results of this study corroborate earlier findings by Heldner et al. (2021) and Hjortdal (2022) stating that accent I facilitates creaky voice more than accent II. Opposing this, figures 13 and 14 display creaky voice being present only in the accent II words. Remedying this point, it is important that the sentence position of the target words is taken into account. Because the ends of phrases often are signalled by creaky voice (Garellek & Keating, 2015), words in sentence-final position feature creaky voice more often. In the elicitation material found in the Appendix, the target words for /o/ and /ʉ/ are found in the beginning and end of phrases. The accent I words are found phrase-initially, while the accent II words are found phrase-finally. It can then be inferred that phrase boundaries have a greater effect on creaky voice than word accents. Thus, though its effect is smaller than phrase boundaries, the results indicate that, in Central Swedish, accent I facilitates creaky voice more than accent II.

Focusing on the method of the study, one could argue that the allophones excluded from the elicitation material, [æ] and [œ], still are interesting to investigate. However, the focus of this thesis is the relationship between vowel dynamics, F0, and word accents; differences between specific vocalic categories is of minimal importance here.

Unfortunately, one pair of words in the elicitation material contained only accent I words. What should have been ‘skyller’ (accent I) versus ‘skylla’ (accent II) became ‘skyller’ versus ‘skyller’. This was not caught until all but one speaker had been recorded. When discovered, the final sentence of the elicitation material was changed from “Den påhittade vännen (/ɛ/, I) som de alltid skyller (/y/, I) på, Fröken Glupsk, har slutat fungera som ursäkt”.

The version found in the appendix is the amended version, in the hopes that the material could be useful for future research.

As for the universality of the potential relationship between F0 movements and formant movements should be investigated. That is to say, whether or not similar correlations can be found in other languages, scandinavian or otherwise, is an avenue for future research which has been relatively underexplored. The effect that tonal contour may have on spectral dynamics also calls into question how research should incorporate prosody more closely with segmental analysis. Should the standard articulation of a vowel in Central Swedish be assumed to be accent I or accent II?

Furthermore, one of the most intriguing aspects of the results were the labial approximant insertions. Release phases for /i/ and /y/ have been attested before (Kuronen, 2000), but not for /ʉː/, /uː/, or /oː/. Surveying this phenomenon more closely is another avenue for future research on Central Swedish specifically.

6. Conclusions:

In summary, this thesis has investigated how formant dynamics and tonal contour interact in Central Swedish, with special focus on diphthongization and word accents. It attempted to answer the research questions of whether the dynamics of F0 affect the dynamics of vowel formants, specifically whether movements of F0 coincide with formant movements, if creaky voice coincides with formant movements, and whether the word accents in Central Swedish facilitate creaky voice or diphthongization to different degrees. In order to answer these questions, recordings of speakers of Central Swedish were analyzed spectrally and temporally.

The compiled results were discussed in relation to the research questions and previous research. Insertions of labial approximants during /ʉː/, /uː/, and /oː/ were found in the studied material. A tentative relationship between F0 and formant movements was found, where the onsets of offglide components in diphthognized vowels often coincided with movements of F0. The interaction between creaky voice and word accents where accent I facilitates creaky voice was corroborated, though no tangible relation between temporal alignment of creaky voice and formant movement was found. Finally, a tentative relationship between word accent and formant movement was found, where accent II facilitates diphthongization for most vowel categories and reduces movements of F1 in /oː/ and /uː/, compared to accent I.

The present study has shown that research into interactions between dynamic F0 and formant movements is warranted, and that phonetic descriptions of Central Swedish still have relatively unexplored territories.

Appendix: Elicitation material

Short story (target words are followed by the target vowel and word accent here, but not in the participants' version):

Flickan, som hette Katja, gick ut i trädgården med en kaffebricka. På brickan fanns några kakor och ett glas saft. Hon hörde vinden vina (/iː/, II) i träden och kom att tänka på alla planer (/aː/, II) de hade hittat på för att göra trädgården ännu finare. Rönnen (/ø/, I) de hade planterat förra året började blomma (/u/, II), och det var så fint att flickan tänkte att hon ville ha flera rönnar (/ø/, II) nästa år. Men rönnar var inte de enda träna (/ɛː/, I) i trädgården, det fanns också björk och asp. I en kruka hade hon sått den bönan (/øː/, II) hon hittat med sina vänner (/ɛ/, II) från gymnastiken. När hon tänkte på dem så fick hon spring i benen (/eː/, I), och hon började träna (/ɛː/, II) på att hjula (/ʉː/, II), men det gick inte så bra. Istället drack hon lite av saften. Den var söt ända bak i gommen (/u/, I) och hon kände sig genast uppfriskad. Hon fixade till benan (/eː/, II), som hade hamnat snett när hon hjulade.

Hennes bror Kalle kom ut och ville ha tillbaka kortleken som han lånat (/oː/, II) ut till henne. Hon gick med på att spela om den. Kalle förlorade eftersom han hela tiden ville syna (/yː/, II) alla korten. Katja kommer hinna (/i/, II) spela många gånger innan han får tillbaka lånet (/oː/, I). Kalle skyller (/y/, I) på att synen (/yː/, I) var så dålig där han satt. Tydligen var skinet (/iː/, I) från solen väldigt störigt. Bönen (/øː/, I) i kyrkan hade inte hjälpt honom vinna den här gången heller.

Snart började de prata om den nya grannen (/a/, II). Den mannen (/a/, I) hade flera hundar (/ʉ/, II), och ena hjulet (/ʉː/, I) på hans bil var helt borta. Sonden (/o/, I) han hade var också spännande. Syskonen hade aldrig sett en sådan innan, men deras mamma hade berättat att vissa äter med sonder (/o/, II). De båda ville gärna ha hund. Planen (/aː/, I) var att skaffa en hona (/uː/, II). Hunden (/ʉ/, I) de ville ha skulle heta Lassi. Kanske hinner (/i/, I) de övertyga mamma innan sommaren är slut. Just nu är de tyvärr i "röda zonen" (/uː/, I), efter att ha ätit upp allt godis innan lördag igen. Den påhittade vännen (/ɛ/, I) som de alltid skyller (/y/, I) på, Fröken Glupsk, har slutat fungera som ursäkt. Den påhittade vännen (ɛ, I) som de brukar skylla (y, II) på, Fröken Glupsk, har slutat fungera som ursäkt

List of sentences (target words are underlined here, but not in the participants' version):

1. Elin mumlade bönen vid ljusbäraren.
2. Anna sådde bönan i en kruka.
3. Emmelie slängde benen i komposten.
4. Johanna har benan till höger.
5. Ulrika tog ut lånet från banken.
6. Sanna skulle låna Marias cykel.
7. Gudrun köpte vinet på systembolaget.
8. Moa hör hur vinden vinar genom träden.
9. Nicole kollade synen i måndags.
10. Ellen ville syna Louise kort.
11. Karin snurrade hjulet en gång till.
12. Hilda lärde sig att hjula på en vecka.
13. Esther sitter i den gröna zonen och väntar.
14. Stina har nu sonat sitt brott.
15. Tina känner till de olika träna i skogen.
16. Alva tycker om att träna på fritiden.
17. Sigrid har hela planen i huvudet.
18. Det finns planer att renovera biblioteket.
19. Räven tycker att rönnen är för hög.
20. Katja har två rönnar i sin trädgård.
21. Maria är den vännen Sanna har haft längst.
22. Att hitta nya vänner är ofta svårt som vuxen.
23. Nour behöver sonden för att få i sig näring.
24. Gabriella har byggt sonder som flugit i rymden.
25. Pia tror att hon hinner till bussen.
26. Ida hade en hinna av svett på pannan.
27. Nathalie går och skyller på sig själv.
28. Emma hade ingen att skylla allting på.
29. Felicia matar hunden med köttbullar.
30. Nikita ogillar hundar som skäller.
31. Tilda har haft ont i gommen sedan helgen.
32. Ika plockar blommor på våren.
33. Isabella tog med mannen till muséet.
34. Kerstin fyllde kannan med vatten.

References:

Baart, J. L. G. (2010). *A field manual of acoustic phonetics*. SIL International.

Boersma, P., Weenink, D., (2024). Praat: doing phonetics by computer [Computer program]. Version 6.4.06, retrieved 25 February 2024 from https://praat.org

Bruce, G. (2007). Components of a prosodic typology of Swedish intonation. In T. Riad & C. Gussenhoven (Ed.), *Volume 1 Typological Studies in Word and Sentence Prosody* (pp. 113-146). Berlin, New York: De Gruyter Mouton. https://doi-org.ludwig.lub.lu.se/10.1515/9783110207569.113

Bruce, Gösta, 'Intonational Prominence in Varieties of Swedish Revisited', in Sun-Ah Jun (ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing* (Oxford, 2005; online edn, Oxford Academic, 1 Feb. 2010), https://doi.org/10.1093/acprof:oso/9780199249633.003.0015, accessed 16 Apr. 2026.

Bruce, G. (1977). *Swedish word accents in sentence perspective*. CWK Gleerup.

Bruce, G. (2010). *Vår fonetiska geografi : om svenskans accenter, melodi och uttal*. Studentlitteratur.

Crystal, D. (2008). *A dictionary of linguistics and phonetics* (6th edition). Wiley. DOI: 10.1002/9781444302776

Eklund, I., & Traunmüller, H. (1997). Comparative study of male and female whispered and phonated versions of the long vowels of Swedish. *Phonetica, 54*(1), 1.

Eliasson, S. (2023). The phonological status of Swedish au and eu: Proposals, evidence, evaluation. *Nordic Journal of Linguistics*, *46*(3), 257–298. doi:10.1017/S0332586522000233

Engstrand, O. (1988). Articulatory correlates of stress and speaking rate in Swedish VCV utterances. *Journal of the Acoustical Society of America, 83*(5), 1863-1875.

Esling, J. H., Moisik, S. R., Benner, A., & Crevier-Buchman, L. (2019). *Voice Quality*. Cambridge University Press.

Ewan, W. G. (1975). Explaining the intrinsic pitch of vowels. *Journal of the Acoustical Society of America, 58*(S1), S40. https://doi.org/10.1121/1.2002115

Garellek, M., Keating, P. (2015). *Phrase-final creak: Articulation, acoustics, and distribution* [PowerPoint presesntation]. LSA 2015. https://linguistics.ucla.edu/people/keating/Garellek-Keating_2015_LSAtalk_creakyvoice.pdf

Gick, B., Wilson, I., & Derrick, D. (2013). *Articulatory phonetics*. Wiley-Blackwell.

Heldner, M., Riad, T., Sundberg, J., Włodarczak, M., & Zora, H. (2021). Pride and prominence. *Working Papers in Linguistics,* 1.

Hjortdal, A., & Lund University, J. F. o. H. a. T. (2022). Creaky voice in South Swedish accent 1. *Proceedings of Fonetik 2022 Phonetic Experimental Research at the Institute of Linguistics (PERILUS) Fonetik 2022, XXXIII*, .

Integritetsskyddsmyndigheten. (16 March 2026). *Behandling av personuppgifter - för forskare*. https://www.imy.se/verksamhet/dataskydd/dataskydd-pa-olika-omraden/forskning/behandling-av-personuppgifter--for-forskare/

Li, C., Al-Tamimi, J., & Wu, Y. (2023, August). Tone as a factor influencing the dynamics of diphthong realizations in Standard Mandarin. In *20th International Congress of Phonetic Sciences (ICPhS)* (pp. 1876-1880). Guarant International.

Li, C., Al-Tamimi, J. (2024). Impact of the tonal factor on diphthong realizations in Standard Mandarin with Generalized Additive Mixed Models. *Proc. Interspeech 2024, 1555-1559*. DOI: 10.21437/Interspeech.2024-303

Lindqvist-Gauffin, J. (1972). A descriptive model of laryngeal articulation in speech. *STL-QPSR*, 13(2-3), 1-9.

Listen Lab. (1 February 2021). *Make your own vowel chart!* [Video]. YouTube. https://www.youtube.com/watch?v=BGW8J4cG0qY&t=2s

Myrberg, S., & Riad, T. (2015). The prosodic hierarchy of Swedish. *Nordic Journal of Linguistics, 38*(2), 115. https://doi.org/10.1017/S0332586515000177

Niebuhr, O. (2004). Intrinsic pitch in opening and closing diphthongs of German. In B. Bel, & I. Marlien (Eds.), *Proceedings of the 2nd International Conference of Speech Prosody, Nara, Japan* (pp. 733-736)

Ohala, John J., Eukel Brian W. (1976). Explaining the intrinsic pitch of vowels. *J. Acoust. Soc. Am*; 60 (S1): S44. https://doi.org/10.1121/1.2003351

Peeters, W. J. M. (1991). *Diphthong dynamics : a cross-linguistic perceptual analysis of temporal patterns in Dutch, English, and German = Diftong dynamica : een contrastieve perceptieve analyse van temporele patronen in het Nederlands, Engels en Duits*. Mondiss.

Pelzer, J. A., & Boersma, P. (2019). Diphthongization in three regional varieties of Swedish. In *Proceedings of the 19th international congress of phonetic sciences* (pp. 1144-1148). Canberra, Australia: Australian Speech Science and Technology Association.

Persson, A. (2025). The acoustic characteristics of Swedish vowels. *Phonetica, 81*(6), 599-643. https://doi.org/10.1515/phon-2024-0011

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.

Riad, T. (1998). Towards a Scandinavian accent typology. In W. Kehrein & R. Wiese (Red.), *Phonology and morphology of the germanic languages* (s. 77–112). De Gruyter. https://doi.org/10.1515/9783110919769.77

Riad, T. (2013). Tonal word accents. *The Phonology of Swedish*. https://doi.org/10.1093/acprof:oso/9780199543571.003.0009

Riad, T. (2013). Vowels. *The Phonology of Swedish*. https://doi.org/10.1093/acprof:oso/9780199543571.003.0002

Ting, C., Sonderegger, M., Clayards, M., & McAuliffe, M. (2025). The crosslinguistic distribution of vowel and consonant intrinsic F0 effects. *Language 101*(1), 1-36. https://dx.doi.org/10.1353/lan.2025.a954227.