Title
Cue Weighting and Enhancement in Systems of Contrast: Vowel Quantity in Norwegian

Author
Block, Aleese Susan

Publication Date
2023

Peer reviewed|Thesis/dissertation

Cue Weighting and Enhancement in Systems of Contrast:
Vowel Quantity in Norwegian

By

ALEESE SUSAN BLOCK
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Linguistics

in the

OFFICE OF GRADUATE STUDIES

at

THE UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

_____
Georgia Zellou, Chair

_____
Santiago Barreda-Castañón

_____
Sverre Stausland Johnsen

Committee in Charge
2023

ABSTRACT

This dissertation aims to provide a detailed description of the primary and secondary cues associated with Norwegian phonological vowel quantity and explore the link between speech production and perception. While vowel duration is the primary, obligatory cue to long-short vowel distinctions in production, the existence and role of secondary cues in Norwegian vowel length is less thoroughly explored. Secondary cues described in Norwegian include longer consonants after short vowels, long vowels that are more peripheral in the vowel space, and long mid vowels that are diphthongized compared to their short counterparts. There also a gap in understanding how listeners use cues in the perception of long and short vowels, particularly knowledge about if and how listeners use secondary cues like vowel quality and postvocalic consonant duration. Experimental studies of the acoustics, hyperarticulation, and perception of long and short vowels were conducted to investigate these topics. The production, enhancement, and perception of long and short vowels as well as theoretical implications thereof are the focus of this dissertation. Three main questions were asked and explored in three experiments: (1) what are the acoustic correlates of Norwegian vowel quantity, (2) how is quantity enhanced by speakers to increase their intelligibility for listeners in Norwegian, and (3) how do listeners use multiple acoustic cues when perceiving the vowel quantity distinction.

In production, Experiment 1 demonstrated that listeners systematically produce long vowels with: (1) longer vowel duration, (2) a different spectral quality than short vowels, and (3) shorter postvocalic consonant duration. Furthermore, long mid vowels were shown to diphthongize in a centralizing direction. Experiment 2 tested how listeners adjusted their speech in clarifying an apparent misunderstanding of their intended vowel length by a simulated interlocutor. Acoustic analysis showed that speakers adjusted their speech by lengthening the duration of long vowels and shortening that of short vowels, producing long-short pairs further apart in the vowel space (except for /ɑ/, and producing consonants after long vowels shorter. Speakers did not, however, enhance spectral movement in long mid vowels. This demonstrates that of the four cues investigated, only three are enhanced for clarity-motivated reasons. Furthermore, there were vowel-specific patterns in that speakers did not enhance the quality difference between long and short /ɑ/. Critically, Experiment 2 also supported accounts that speakers will adjust their articulations in a targeted way that is aimed at eliminating perceptual confusability of phonological contrasts. In perception, Experiment 3 tested how listeners used the acoustic correlates of quantity found in Experiments 1 and 2 and whether there would be vowel-specific perceptual strategies. Analysis showed that listeners did have vowel-specific patterns in their perception of long and short vowels. While vowel duration was used for all six pairs, vowel quality was not reliably used for long and short /ɑ/ and postvocalic consonant duration was not reliably used for long and short /u/.

The results of the experiments outlined in this dissertation suggest that (1) speakers produce long and short vowels with multiple acoustic correlates, (2) these acoustic qualities are enhanced in targeted ways to increase intelligibility of speech, and (3) listeners use secondary cues in perception with both cue- and vowel-specific patterns. The theoretical implications of these findings were further discussed in relation to cue weighting, clear speech, the production-perception link, and cross-linguistic patterns.

Dedicated to my niece, Lilly Eggers. Through my efforts to be the role model you deserve I have become a better version of myself.

Acknowledgements

Writing a dissertation is hard (believe it or not). There's a saying about it taking a village to raise a baby. I think that this also applies to raising a Ph.D., there is no way I could've pulled this off alone. I am grateful for the following people for their support during the completion of this dissertation and everything they added to the journey.

My deepest gratitude goes to my advisor, Georgia Zellou. From my first day in my program, Georgia's encouragement, perspectives, and endless support have motivated me to get to the point I am at today. I am also indebted to my other committee members for everything they have done. I thank Sverre Stausland Johnsen. If it were not for his interest in my research, support, and willingness to host me in Oslo, I would not have been able to get the data and the experiences I needed to complete this dissertation. I also want to thank Santiago Barreda-Castañón for his big-picture perspectives and help with my statistical skills. Additionally, I am thankful to our Graduate Program Coordinator, Stephanie Fallas. If it were not for her constant support and humor, I think I (and many other graduate students) may have lost my mind. Lastly, I want to thank other mentors I have had along the way, including Anne Pycha at the University of Wisconsin-Milwaukee, who showed me the early years of my university education that I was worthy and capable of pursuing a Ph.D.

I want to thank everyone who partook in my experiments, giving their time and knowledge and receiving nothing in return. Their generosity made this dissertation possible. And also to the American-Scandinavian Foundation for funding my data collection during the 2021-2022 school year.

My fellow graduate students have also contributed greatly to my growth throughout my program. Thank you especially to Chloe Brotherton, Tyler Méndez Kline, and Kristin Predeck for the support, laughs, late nights (both working and not), and your friendship. Without your comradery, getting through these five years would've been like walking through a maze alone.

Importantly, I am thankful to my family. I thank my dad, Steve, for always supporting every endeavor of mine, even if he didn't always understand what I was talking about. His constant certainty that I can accomplish anything that I put my mind to has helped me overcome many obstacles both big and small. I thank my big sister Katie for being my first inspiration to pursue academia. She showed me that even I could one day have a "fancy shmancy title". I thank my mom, Sue, for letting me live with her during the worst of the covid pandemic and constantly reminding me that the hard times are only temporary and that soon I would be able to continue with my research plans and my big move to Norway. Thank you to my sisters Ashley and Abby for always being a phone call away and keeping life interesting.

I thank my partner Tomas and his family. Moving to a different country to write my dissertation has been one of the greatest and most challenging experiences of my life. Thank you to Tomas for his support, love, patience, and many delicious homecooked meals. And for answering my hundreds of questions about Norwegian and helping with my experiments. Lastly, thank you to the Ødegård family for welcoming me in and encouraging me along the way. Because of you, Norway has felt like home.

Lastly, I want to thank my cat Åsa for being the world's best distraction whenever I needed a break. Your big brother Oliver would've loved you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 - INTRODUCTION

1. STATEMENT OF THE PROBLEM

This dissertation explores the link between phonetic variation and systems of phonological contrast. Liberman et al. (1967)'s "lack of invariance" problem outlines the fact phonemic category membership is rarely signaled by a single acoustic cue, alluding to the inherently robust and multidimensional nature of phonological contrast. When using the term "cue", we will work off the definition: "[…] a perceptual cue might be considered any information that systematically influences listeners' perception of a contrast, while a cue in production could encompass anything that varies systematically across members of a contrast" (Schertz & Clare, 2020, p. 2). This idea of multiple cues to define a contrast is explored by Enhancement Theory, which further describes the function of redundant, enhancing secondary cues in creating more salient contrasts for both speakers and listeners (Stevens & Keyser, 1989; Kingston & Diehl, 1994; Keyser & Stevens, 2006). Kingston and Diehl's (1994) Duration Ratio Hypothesis is another example of this. In this account, the relative durations of vowels and consonants in VC clusters, specifically when one has a quantity contrast, are mutually enhancing cross-linguistically. Despite several theories dealing with multiple cues in phonological contrast, the exact role that phonetic variation plays as potentially enhancing secondary cues is yet to be understood fully. Thus, one of the central goals of this dissertation is to further explore how secondary acoustic cues illustrate the complexity of seemingly simple phonological contrasts, pointing to the rich nature of contrast in general.

Furthermore, this dissertation explores the relationship between speech production and perception. Speech production and perception have traditionally been studied independently but understanding their relationship and where they converge and diverge helps to create a comprehensive model of representation and linguistic knowledge (Casserly & Pisoni, 2010; Schertz & Clare, 2020). When considering the relationship between these two modalities, we should consider the directionality thereof. For example, if production were guided by perception, speakers make use of cues in production as they are used in perception with the goal of guiding a listener (Beddor et al., 2018). Or if perception were guided by production, listeners' perceptual judgements could be based on the same metric they use as speakers when producing a contrast (Newman, 2003). Conclusions on how cues in production and perception are correlated are complex. Schertz and Clare (2020) state that no correlations between production and perception have been found in any "standard" cue-weighting study, defined as a study looking directly at cues for a phonological contrast as compared directly in production and perception. However, significant correlations have been found in studies focusing on contextualizing cues. For example, Zellou (2017) found a correlation between degree of nasalization in production and sensitivity in perception. Yet, this is an area which needs to be investigated further to be understood. Therefore, using acoustic correlates of Norwegian vowel quantity as an example, this dissertation will also examine the link between cues as used in production and in perception. Specifically, do we see that the acoustic correlates of contrasts are mirrored in what listeners use in perception? And if there is a mismatch, what does that mean for the overall relationship between production and perception?

These larger questions will be addressed through the lens of primary and secondary acoustic cues in the production and perception of Norwegian vowel quantity. Contrastive vowel quantity is when the duration of a vowel conveys lexical meaning: whether a vowel is long or short encodes a different word. While vowel duration is considered the primary, obligatory cue to vowel quantity, there are a number of secondary cues that have been attested in production including f0, spectral characteristics, and postvocalic consonant duration.

Descriptions of vowel quantity in Norwegian vary in terms of what is or is not included in this contrast. One group of descriptions take a simplified stance that the only temporal cues were correlated with quantity in production, namely vowel duration (Behne et al., 1996) and postvocalic duration (van Dommelen, 1999). Other accounts include differences in vowel quality, describing short vowels as being more central than long vowels (Kristoffersen, 2000, p. 16); this is in line with cross-linguistic descriptions of vowel quantity (Maddieson, 1984, p. 129-130). The latter stance that long-short pairs have quality differences has been long accepted within by Norwegian scholars.

Empirical studies on the role of primary and secondary cues in the perception of Norwegian have been small, sparse, and rather unclear in their conclusions. Both Nylund and Behne (1996) and Behne and Nylund (2003) examined the role of spectral and durational information in quantity perception and provided some preliminary evidence that listeners might utilize vowel quality in perceiving quantity. However, this was only shown to occur for certain vowels within the subset used in the experiments and the exact size of the effect was not discussed at length. Van Dommelen demonstrated that the duration of the postvocalic consonant had an effect on the categorization of long and short vowels,

specifically that changing the postvocalic consonant duration moved the location of the perceptual boundary between quantities. It is worth noting that the roles of vowel quality and postvocalic consonant duration have been investigated separately in research on Norwegian, leaving a gap in the literature in terms of having complete and comprehensive account.

The situation of vowel quantity in Norwegian raises many questions about the relationship between phonetic variation and phonological contrast. Hence, there are several aims of this dissertation:

With respect to the production of phonological vowel quantity, this dissertation will serve two main purposes. First, this dissertation will be one of the most comprehensive acoustic descriptions of how vowel quantity is produced in Norwegian. What are the exact acoustic qualities that Norwegian speakers produce? Are there vowel-specific patterns? In Chapter 2, this dissertation investigates the ways in which vowel quality, vowel duration, postvocalic consonant duration, and vowel inherent spectral change (VISC) signal quantity in perception. The second purpose of the production aspect of this dissertation is to provide the first study on how Norwegian vowel quantity is enhanced in clear speech production, which is investigated in Chapter 3. Examining the enhancement strategies of speakers as it relates to vowel length in Norwegian can give insight into how the contrast works in the language. Furthermore, this provides an opportunity to examine how speakers adjust their language in local, contrast-oriented ways aimed at getting minimizing sources of perceptual ambiguity. Lastly, the study will investigate any cue- and vowel-specific patterns that may arise.

With respect to perception, how do listeners utilize both temporal and spectral cues in perceiving vowel quantity? How do listeners adapt their weighting of various acoustic cues when one or more cues are no longer informative? Chapter 4 will explore how listeners weight vowel quality, vowel duration, and postvocalic consonant duration in identifying phonemic quantity. In addition to if and how much listeners rely on each cue, vowel-specific perceptual strategies will be investigated as well.

Finally, the implications of this research for our understanding of the role of sub-phonemic information in the phonetic grammar of Norwegian are explored. For example, if secondary cues like vowel quality and the duration of the postvocalic consonant are systematically associated with the production and perception of vowel quantity, should we consider them to be part of the phonological representation of these segments? And if there is a difference in how these cues are integrated into phonological quantity between production and perception, how can we reconcile this? The theoretical implications of these questions are an important focus of this dissertation as well.

2. LITERATURE REVIEW

2.1 Vowel quantity

2.1.1 Production

The primary, obligatory acoustic cue in *segment* quantity is duration. Critically, durational differences which signal quantity are said to be phonemic when a language uses the long and short forms of a vowel or consonant (geminate vs. singleton) to encode lexical

distinctions. For example, Japanese has both types of contrasts as seen in (1). Vowel quantity is demonstrated by the contrast of a short vowel in (a) and long vowel in (b); gemination is demonstrated by the contrast of a singleton (short) consonant in (c) and a geminate (long) consonant in (d).

(1) Japanese quantity contrasts (Tsujimura, 2007)

a. [su] “vinegar”

b. [suː] “inhale”

c. [saka] “hill”

d. [sakːa] “author”

In addition to more common long-short distinctions in quantity, some languages have three-way distinctions. For example, Estonian has short, long, and extra-long vowels as seen in (2).

(2) Estonian vowel quantity contrasts (Lippus, 2011)

a. [sata] “hundred”

b. [saːta] “send (imperative)”

c. [saːːta] “get (infinitive)”

It is important to note that segmental duration is usually not the sole cue present in production that marks quantity. It has been well established that vowel quality differences exist in long/short vowel pairs in a number of languages. Specifically, short vowels tend

to be realized more centrally in the vowel space than long vowels. Maddieson (1984, p. 129-130) surveyed 331 languages and found 56 languages with vowel quantity distinctions. Of these 56 languages, 17 languages were shown to have centralization of the short vowels in some or all pairs of vowels. Table 1.1 demonstrates the distribution of these quality contrasts within the vowel space.

Table 1.1: Breakdown of 56 languages (in Maddieson, 1984, p. 129-130) that have quality differences between long and short vowels in the given area of the vowel space

Vowel Quality	**Attested Languages**	**Difference in Quality**	**Percentage**
High Front	40	17	42.5
High Central	2	0	0
High Back	37	10	27
Mid Front	27	16	59.2
Mid Central	2	1	50
Mid Back	28	14	50
Low Front	7	1	14.3
Low Central	31	0	0
Low Back	4	0	0

Differences in f0 have been attested in quantity contrasts too. The most commonly cited language for f0 differences between long and short vowels is Japanese. In Japanese, short vowels are typically produced with a static f0 while long vowels are produced with

a falling f0 (Kinoshita et al., 2002). While differences in quality between long and short vowels is relatively common across languages with this phonemic contrast, f0 is significantly less common.

Acoustic differences are not always limited to the quantitative segment; oftentimes, adjacent segments undergo some degree of compensatory lengthening or shortening. For example, consonants following long vowels are typically produced shorter than those following short vowels, and this can be seen in languages such as Icelandic and Swedish (Pind, 1996; Behne et al., 1999). This follows Kingston and Diehl's (1994) Duration Ratio Hypothesis, which states that the duration of a vowel and the following consonant are mutually enhancing, regardless of which segment is quantitative.

From here, it is abundantly clear that phonological quantity is not marked solely by segmental duration. Rather, multiple enhancing, secondary cues help to signal long and short vowels. Cross-linguistic research on the contrast between long and short vowels is important to understanding the multidimensional nature of not only vowel quantity contrasts but phonemic contrasts in general. Therefore, an important question to continue investigating is, what are these multidimensional acoustic properties of vowel quantity and how do these manifest cross-linguistically?

2.1.2 Perception

In the pursuit of understanding how vowel quantity operates both within and across languages, acoustic correlates found in production are naturally an important component. However, we know that speech does not exist in a vacuum and is produced with the intention of being understood by a listener. Thus, another important aspect to consider is

which acoustic cues are informative for listeners when determining the difference between long and short vowels. As with studies in production, comprehensive and cross-linguistic examinations of the cues that listeners use in perception are needed to shed light on the complex nature of quantity contrasts. Another interesting question that requires further investigation is how the qualities that correlate with quantity in production are used by listeners in perception. Specifically, do we see patterns in production mirrored in perception? And what does the answer we receive tell us about the nature of the relationship between production and perception as it pertains to this phonological contrast and linguistic systems as a whole?

Previous work has looked at the acoustic cues related to the perception of vowel quantity in several languages. Specifically, work has attempted to uncover which of the secondary cues described in the previous section are most salient for listeners during word comprehension. While the most salient cue for listeners has been established to be vowel duration, secondary cues such as segmental context (Tranmüller & Krull, 2003), dynamic f0 (Lehiste, 1976; van Dommelen, 1993), and spectral differences between long and short vowels (Abramson & Ren, 1990; Sendelmeier, 1981) have been shown to impact listeners' judgement of vowel quantity.

Early investigations of the perception of dynamic f0 found that listeners tended to perceive vowels with a dynamic f0 as longer than vowel with a level f0 (Lehiste, 1976; Pisoni, 1976; Wang et al., 1976). Wang et al. (1976) also found that vowels with a rising f0 contour were perceived as longer than those with a falling f0 and vowels with a falling fundamental frequency, in turn, were perceived as longer than those with a level f0. However, later studies on the topic either found that an increase in perceived vowel

duration due to f0 was context dependent (van Dommelen, 1993) or failed to replicate this effect altogether (Rosen, 1977). It is also worth noting that studies that found perceptual lengthening due to a dynamic f0 have been done primarily on American English, which does not have a vowel length contrast, while those that did not consistently find an effect were done on language with phonemic vowel quantity (Swedish: Rosen, 1977; German: van Dommelen, 1993).

Multiple studies have explored the role of vowel quality in the perception of vowel quantity across various languages. Some earlier literature asserted that spectral vowel characteristics were not used in quantity perception (e.g., Garnes, 1976); however, these studies used long-short vowel pairs that did not display significant spectral differences in their production. Other studies specifically targeted testing vowels where long and short phonemes were spectrally dissimilar. Pind (1996) tested the use of spectral characteristics in the perception of Icelandic vowel quantity. Using long and short /ɑ/ and /ɛ/, stimuli manipulated along both vowel duration and spectral quality were created. Listeners were presented with the stimuli and asked to choose from two given words, which they had heard, categorizing the vowel as long or short. Pind found that spectral factors can be of “decisive importance” during the perception of Icelandic vowel quantity.

Lehnert-LeHouillier (2010) conducted a cross-linguistic examination of the role of secondary acoustic cues in the perception of vowel quantity in three quantitative languages: Thai, Japanese, and German. They found that for listeners of all three languages, vowel duration was an important cue for the vowel quantity identification, and the duration of the phonemic boundary between long and short vowels differed between languages. F0 was found to influence the perceived vowel duration only for Japanese

listeners whose native language associated long vowels with a dynamic (i.e., falling) f0. This finding is similar to Behne et al. (1999), who had also found that Japanese listeners interpreted a falling f0 as signaling a long vowel. Lastly, all listeners were influence by spectral cues in their judgement of vowel quantity. The findings of this study bring about the point that some cues (e.g., f0) are seen to signal vowel quantity in production for some languages but not others, and this can be mirrored in perception via the emergence of language-specific perceptual patterns.

Lippus et al. (2013) looked at the role of duration, pitch, and vowel quality in the perception of the three-way (short, long, and extra-long) Estonian vowel quantity distinction. In the case of Estonian, the researchers point out that in the case of languages that have a three-way quantity distinction (e.g., relatively common in Finno-Ugric languages), it is extremely common for the contrast to be marked by at least one prosodic feature to increase distinctiveness. They found that vowel quality was important in distinguishing short and long/extra-long vowels; this was because the largest difference in vowel quantity was here, with short vowels being more centralized than long, but no significant spectral difference occurring between long and extra-long vowels. The researchers found that f0 was important for listeners in perception while distinguishing between long and extra-long vowels; this also makes sense, as there is a clear difference in pitch and contour in the production of long and extra-long vowels that does not necessarily exist with short vowels. Finally, they found that the duration of the coda consonant was an important factor in identifying all three quantities. This study displayed that it is not only features directly on the segment that aid in quantity identification, but also acoustic qualities of adjacent segments as well.

Taken together, the results of these studies support the robustness and multi-dimensionality of phonemic vowel quantity. Not only do we see a number of acoustic correlates of quantity in production, these primary and secondary cues are informative for listeners in perceiving. Furthermore, these studies show language-specific patterns in terms of what cues are included in signaling vowel quantity, giving merit to examine languages closely to learn what that particular language uses.

2.2 Multiple acoustic cues perception

2.2.1 Cue weighting

A single acoustic dimension is rarely sufficient to define phonological category membership, a classic illustration of the "lack of invariance" issue outlined by Liberman et al. (1967). Enhancement Theory begins with the underlying assumption that in any given language, there is a set of contrasts signaled by acoustic cues that are often enhanced to be more robust and salient for listeners. Enhancement of a phonological contrast by covariation with another feature is common cross-linguistically (Stevens & Keyser, 1989; Kingston & Diehl, 1994; Keyser & Stevens, 2006).

An important question to ask is: how do listeners handle multiple acoustic cues covarying with a single contrast in the speech signal? In terms of multiple cues signaling a single contrast, some cues in production are more strongly correlated with category membership than others and this is mirrored in how much listeners use these cues (Abramson & Lisker, 1984; Idemaru & Holt, 2011). When we talk about some acoustic cues being more informative for listeners, this refers to the phenomenon where some

cues are strongly correlated with listener categorization responses while others are not as predictive of perceived sound category. The fact that some acoustic dimensions play a greater role in determining the perceptual identity of a phoneme is referred to as cue weighting; the ability of a listener to integrate and weight acoustic information across different dimensions is critical for speech perception (Holt & Lotto, 2006).

A common example used to illustrate cue weighting is the way listeners handle spectral and temporal acoustic cues in the perception of English tense-lax vowel distinctions. It has been established that tense vowels are systematically produced with longer durations than lax vowels. Hillenbrand et al. (2000) tested listeners in the identification of twelve American English vowels in an /hVd/ context after manipulating vowels along both the duration and spectral dimensions. They found that while listeners utilized both spectral and durational information in their identification of vowels, they relied more heavily on spectral cues than vowel duration.

However, cue weighting is not always uniform across contexts, contrasts, or individuals. Clayards (2018) compared individuals' cue weights within and across five different contrasts. For each contrast, stimuli were manipulated along a primary and a secondary acoustic dimension (see Table 1.2). Listeners completed a two-alternative forced choice task for four out of the five sets of minimal pairs. They found that the way in which listeners weighted primary and secondary cues in relation to one another differed across contrasts; for some contrasts, the weight of primary and secondary cues was positively correlated while for others they were negatively correlated. Furthermore, Clayards found much individual variation in how cues were weighted.

Table 1.2: Minimal pairs and the primary and secondary acoustic cue manipulated in Clayards (2018)

Minimal Pair	Primary	Secondary
bet-bat	Formant frequency	Vowel duration
bog-dog	Vowel transition	Release burst
dear-tear	VOT	Onset f0
Luce-lose	Duration ratio	Vowel Transition
sock-shock	Frication noise	Vowel transition

One central question within research on cue weighting is: what determines the relative cue weighting of different acoustic dimensions in the speech signal? Holt and Lotto (2006) state than an adaptive listener would weight dimensions based on experience over time with the acoustic environment and that we could predict weighting functions for speech perception if we knew how acoustic dimensions co-varied with phonetic contrasts in a listener's experience. They then go on to describe four variables that influence how cues are weighted, with two relating to the distributional characteristics of acoustic information and being language specific, and the other two not. Here, I will address the first two points. First, cues can differ in their informativeness in category identity, and this informativity can be determined by the distinctiveness of the distributions of categories along the acoustic dimension: if the category distributions are more distinct along one dimension, it is more informative. For example, VOT in American English are quite reliably different across categories and is therefore a very informative cue for stop voicing contrasts (Lisker and Abramson, 1964; Lotto & Holt, 2006).

The second factor in how cues are weighted comes from a perceptual learning perspective and is variance; the auditory system seems to be especially sensitive to dimensions that are varying (Lotto & Holt, 2006). This concept was demonstrated with non-speech sounds, where components with greater relative variance were more heavily perceptually weighted (Lutfi & Doherty, 1994). Yet, there is evidence that the relationship between within-category and between-category variance may be a determinant of perceptual weight. While greater distinctiveness between category can lead to greater informativity for a particular cue, large within-category variance can decrease the informativeness of a dimension by potentially creating distributional overlap (Holt & Lotto, 2006).

One must remember that these two factors in how listeners weight cues should be considered language, dialect, and perhaps even listener specific. For example, while aspiration might be a strong cue to voicing in English, it does not carry the same weight for Hindi stop consonant categories because it is strongly correlated with the aspirated-unaspirated distinction (Benguerel & Bhatia, 1980). Therefore, when examining the weighting of multiple cues for a contrast in a language, it is worthwhile to closely examine how that contrast is produced in that language. The link between distinctiveness and variation in the perception of a cue and how those cues are subsequently used in perception in a corner of speech perception that has not been robustly explored and warrants more attention in further research.

2.2.2 Flexibility in speech perception

Idemaru and Holt (2011) called phonetic category restructuring based on category internal information "dimension-based statistical learning", where listeners will dynamically adjust the use of various acoustic dimensions that define phonetic categories. Liu and Holt (2015) examined this paradigm in native English listeners' perception of vowels, specifically looking at the weighting of the primary cue of spectral quality and the secondary cue of vowel duration. They found that while listeners primarily (at baseline) used spectral quality with vowel duration being secondary, they flexibly down-weighted their use of vowel duration when exposed to an artificial accent that deviated from English norms (i.e., filtered out counter-productive acoustic information.

There is evidence that listeners are able to also use distributional information in the input to learn which dimensions are more reliable overall; specifically, listeners will increase their use of a secondary dimension when the most reliable dimension is no longer informative. Kim et al. (2020) examined listener weighting of vowel quality (primary) and duration (secondary) in perceiving American English tense-lax vowel distinctions. After establishing a baseline cue weighting for listeners, the researchers exposed participants to Korean-, Italian-, or Mandarin-accented English vowels; these stimuli were manipulated to deplete the informativeness of spectral information but enhance the vowel duration. Afterwards, listeners completed an identification task with the accented tokens. The researchers found that participants flexibly down-weighted spectral information and up-weighted vowel duration for the accented tokens, demonstrating the dynamic adjustment of cue weights described by Idemaru and Holt (2011). Understanding how listeners handle speech that deviates from established native

norms is useful. For example, speech that deviates from native language norms, requiring enhancement by non-primary acoustic dimensions, is not uncommon: non-native pronunciations of English front vowel contrasts (e.g., (/i/-/ɪ/) tend to have exaggerated vowel durational differences and spectral differences that are less pronounced (Escudero, Benders, & Lipski, 2009).

While most studies have focused on this phenomenon on a group level, a smaller body of work has examined individual differences as well. For example, Schertz et al. (2016) found large individual differences in individual cue re-weighting strategies in listeners presented with foreign accented words. The researchers tested Korean speakers' productions and cue weighting when perceiving Korean (L1) and English (L2) stop voicing contrasts. They found that while participants reliably used both VOT and f0 in their productions, the way in which participants adjusted their cue weighting between hearing L1 and L2 tokens varied widely across individuals, with no prevailing pattern emerging.

Research on the flexibility in speech perception adds nuance and depth to our understanding of cue weighting. While early research on cue weighting attempted to find a clear-cut and widely applicable pattern, more recent research has embraced the concept of cue weighting being language, context, and even individually specific. Research on flexibility in speech perception offers perspective on how cue weighting within an individual can change based on distributional properties of the input, highlighting how dynamically adaptive the listener really is. Furthermore, what we can learn from how listeners adapt their cue weighting strategies can offer insight into cue informativity and how various acoustic cues can be used to signal phonological contrasts.

2.3 Enhancement in production

As discussed above, we know that phonological contrasts are often marked by multiple acoustic cues in production. These cues are not only useful in signaling contrast in the production of speech but are used by listeners in perception. Listeners are adaptive and speech perception is dynamic; listeners are able to take the most informative acoustic cue and weight it heaviest when distinguishing between phonemes and this process is not static. As outlined in “dimension-based statistical learning”, when one cue that was once informative is no longer so, listeners are able to up-weight other cues in order to successfully decode the speech signal. Given the dynamic nature of speech perception, it is worth investigating if this adaptivity extends to speech production too. When there are multiple cues to produce to signal a contrast, are speakers able to adjust their speech across communicative contexts to better signal a phonological contrast? How can speakers enhance cues to make speech more intelligible? Research investigating the acoustic characteristics of various types of speech and adaptive adjustments is discussed below.

2.3.1 H&H Theory and clear speech

Lindblom’s H&H (hypo- and hyperarticulation) Theory is an account of variation based upon communicative context. H&H Theory illustrates that speech is adaptive in that speakers typically change their performance according to communicative and situational demands. Furthermore, speech production exists on a continuum between hypo- and

hyperspeech, each marked by their own characteristics, and this is regulated by balancing listener- and speaker-oriented forces. Hypospeech occurs when the communicative context favors the listener, and the speaker can reduce articulatory effort and produced reduced speech. On the other hand, when the communicative context is such that the listener may have difficulty receiving the intended message, the speaker will produce clearer, hyperarticulated speech. Situations that can elicit hyperspeech vary greatly and can include noisy external environments or a communicative participant who is an L2 speaker of the language used.

There are a handful of characteristics deemed to be global, in that they persist across talkers and languages. For example, clear speech is often produced 5 to 8 dB greater than conversational speech (Picheny et al., 1986), similar to speech produced in noise (Bond et al., 1989) or when shouted (Rostolland, 1982). Furthermore, clear speech has been found to be slower, with speaking rates of 90 to 100 wpm compared to 160 to 205 wpm in conversational speech (Picheny et al., 1986). Similarly, Bradlow (2002) found in their one male and one female talker that sentence duration increases between 51% and 116% when changing into clear speech. Possibly contributing to increased sentence durations, it has been reported an increase in the number and duration of pauses; here, a pause was defined as any silent interval between words greater than 10 ms excluding silence before word-initial plosives (Picheny, 1986; Krause & Braida, 2004; Bradlow (2002). In addition to loudness and duration, clear speech also often has a higher f0 and a larger range, suggesting more laryngeal tension, similar to speech produced in noise (Bond et al., 1989; Summers et al., 1988); however, these changes are not necessarily consistent across talkers (Picheny et al., 1986; Krause & Braida, 2004).

There has been evidence for more segment-focused effect in clear speech. For example, vowels in clear speech are often produced with an expanded vowel space and increased duration (Chen, 1980; Picheny et al., 1986; Moon & Lindblom, 1994; Bradlow, 2002; Krause & Braida, 2004). While Bradlow (2002) found similar degrees of vowel space expansion across the vowel inventory for English and Spanish speakers, Krause and Braida (2004) found differences based on vowel tenseness (i.e., only tense vowels expanded). In addition to adjustments to the vowel space, vowel duration differences have been found to be enhanced in a way correlated to language-specific phonological structure. Smiljanic and Bradlow (2008) examined clear speech production of English tense and lax vowels, English vowels before voiced and voiceless stops, and Croatian long and short vowels. They found that for the English tense-lax distinction, there were no significant changes in durational contrast yet in Croatian, where duration is central to the long-short vowel contrast, there were changes in durational differences. This suggests that while vowel space expansion is a common hallmark of clear speech, temporal cues such as vowel duration in the case of languages with vowel quantity may also be commonplace.

Taken together, the results from previous studies on segmental contrast enhancement in clear speech indicate that hyperarticulated speech tends to enhance acoustic distance between phonological categories and this may be language-specific (Kang & Guion, 2008). Therefore, the enhancements made in clear speech could be a viable route by which to investigate the cues that are central to a phonological contrast.

2.3.2 Error resolution

One line of inquiry has examined clear speech in a specific communicative context: when a listener mishears a speaker. Specifically, are speakers able to dynamically adjust their productions to suit specific communicative challenges? While we might be able to see a general pattern in regular clear speech, do speakers also make adjustments that are specifically targeted at sources of phonological confusion? Targeted adaptation accounts propose that speakers dynamically adjust their speech to address local communicative issues (Lindblom, 1990; Buz et al., 2016) and speakers making phonetic modifications while clarifying misheard speech has been documented by several studies. Ohala (1994) found durational increases in vowels and voiceless stop consonants and Oviatt et al. (1998) found that speakers globally increased duration of speech segments and pauses as well as exaggerated intonational contours.

Ohala (1994) also examined whether durational differences were larger on the specific segment that had been misunderstood compared to surrounding segments. For example, would the VOT of a voiceless stop (e.g., in pit) be increased more if the interlocutor had misunderstood the voiceless stop as voiced (e.g., as bit) compared to if the interlocutor had misunderstood an adjacent vowel (e.g., as pat). Ohala found no significant differences. However, further work has uncovered differences in “global” and “focal” hyperarticulation occurring during error correction. Oviatt et al. (1998) analyzed the speech of participants correcting a simulated speech recognizer that produced two types of errors: (1) general error, where the system replied with “???”, or (2) substitution, where the system guessed the wrong word or phrase. While durational increases were

reported across corrections for both types of errors, they were larger for focal error correction.

Schertz (2013) expanded on these findings and recorded speakers producing speech toward what they believed was an automatic speech recognition system. Target words had either a voiced or voiceless onset plosive. Similar to Oviatt et al. (1998), the system either showed a general error or a more specific error; specific errors included mistakes in voicing, place of articulation, or manner of articulation. Schertz (2013) found that VOT to mark voicing was hyperarticulated when the interlocutor misheard the voicing of the onset plosive specifically; there was no hyperarticulation of the onset VOT when the mistake was general or in the place or manner of articulation. Additionally, hyperarticulation was only on the VOT: neither overall amplitude nor overall word duration was hyperarticulated.

Buz et al. (2016) introduced the Adaptive Speaker Framework, another targeted adaptation account. They investigated how speakers adapt their productions when feedback from their interlocutors suggests that previous productions might have been perceptually confusable. Through a pseudo-interactive task, participants gave instructions to a simulated partner with naturalistic response times. The researchers manipulated whether the target word, which contained a voiceless plosive in onset position (e.g., pit), occurred in the presence of a competitor with a voiced onset plosive (e.g., bit) or an unrelated word (e.g., food). They found that participants hyperarticulated VOT specifically in the presence of a voiced competitor, but not in the presence of an unrelated word. It is important to note that these results occurred in the absence of explicit clarification requests, with the hyperarticulation of VOT suggesting that listeners

preemptively hyperarticulated VOT in situations where there may be perceptual confusability. While other descriptions of targeted speaker adaptation claim that speakers make adjustments based on real-time communicative difficulties, this account claims that these adaptations are segmentally-targeted specifically to the phonological source of confusion.

Cohn et al., (2022) examined whether adjustments made in speech might not just be contrast- or context-specific, but also specific to a type of interlocuter. They tested whether speakers had targeted error correction strategies for voice-AI and human interlocutors. Across two studies with varying rates of comprehension errors, they found that speakers did indeed have differences in strategies between interlocutor types: speakers produced louder speech with a lower f0 and smaller f0 range when correcting errors from a voice-AI system than an apparent human interlocutor. Speakers also produced more vowel hyperarticulation with the voice-AI interlocutor as well. These findings add support to the account that speakers are able to adjust their articulations in very targeted manners not only accounting for the nature of a mistake or phonological contrast, but also for the type of interlocutor with which they are interacting.

Together these studies illustrate how speakers hyperarticulate in the specific context of an interlocutor misunderstanding the intended linguistic message and that hyperarticulation can be a “targeted and flexible adaptation rather than a generalized and stable mode of speaking” (Stent et al., 2008, p. 163).

3. TARGET LANGUAGE: NORWEGIAN

3.1 Vowel inventory

According to Kristoffersen (2000[1], p. 13), the set of surface vowels that can exist in stressed syllables in Urban East Norwegian (UEN) is:

Long: [iː, yː, ʉː, uː, eː, øː, oː, ɑː, æː]
Short: [i, y, ʉ, u, ɛ, œ, ɔ, ɑ, æ]

In addition to the monophthongs described above, there are six diphthongs found in Norwegian (as described in Kristoffersen, 2000, p. 19):

Common: [æj, œj, æw]
Marginal: [ɔj, ʉj, aɪ]

Here, the marginal diphthongs are those that only appear in a small, mostly borrowed, number of Norwegian words.

[1] It should be noted that the descriptions given in Kristoffersen (2000) refer to the dialect of Urban Eastern Norwegian (UEN).

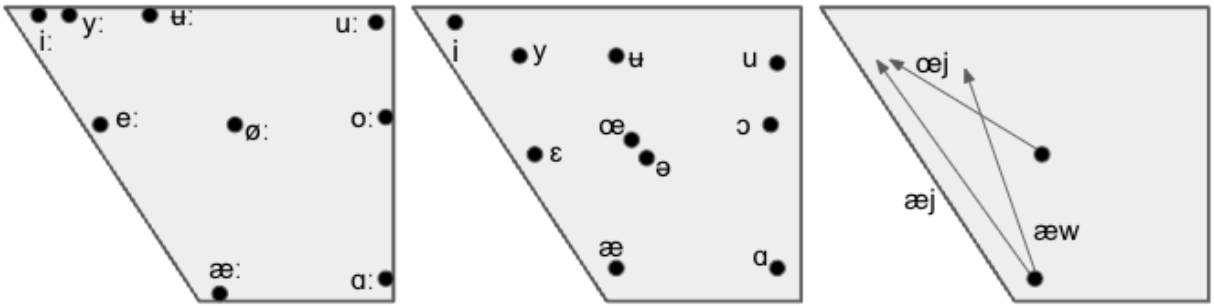

Figure 1.1: Placement of long monophthongs (left), short monophthongs (middle), and diphthongs (right) in the vowel space (based on Kristoffersen, 2000, p. 17)

3.2 Vowel quantity

To understand the synchronic state of quantity in Norwegian, we can begin with a diachronic perspective. All Germanic languages historically had quantity contrasts for both consonants (geminate-singleton) and vowels (long-short), but various processes such as open syllable lengthening and degemination occurred during the medieval period. These changes, referred to as the Germanic quantity shift, were not uniform across the language family: for example, while most varieties of German underwent degemination, Swiss and Bavarian varieties did not (Page, 2020). While it was happening between dialects within-language, variation was even more pertinent between-language. Therefore, some Germanic languages retained vowel quantity, and some retained both (Davis, 2011; Page 2020).

Languages which underwent open syllable lengthening via vowel lengthening but not gemination include Faroese, Swedish, and Norwegian (Kristoffersen, 2011; Page, 2020). Norwegian has contrastive vowel quantity, such that differences in vowel duration are phonemic and distinguish words as those in (3):

(3) Norwegian quantity pairs (Kristoffersen, 2000)

a. [tɑːk] “roof”

b. [tɑk] “thank you”

c. [viːn] “wine”

d. [vin] “win”

Generally long vowels have been reported to be anywhere from 1.4 to 3.3 times longer than short vowels, depending on the source (see Table 1.3). A traditional view of this length contrast asserts that the only difference between quantitative vowel pairs is the duration of the vowel. Behne et al., (1996) recorded 12 native speakers of Norwegian producing long and short /i, o, ɑ/ in real monosyllabic words with either /g/ or /k/ codas for a total of twelve words; the researchers did not use minimal pairs in this experiment. From these recordings, they took three measurements: (1) vowel duration, (2) postvocalic consonant duration, and (3) F1 and F2 values from the midpoint of the steady state of the vowel. The found that while there was an effect of vowel quality on duration in that low vowel /ɑ/ was generally longer with a slightly larger difference in duration in the long-short pair, there was not an effect of quantity on quality.

On the other hand, changes in vowel quality have been cited in other descriptions: specifically, long vowels are more peripheral in the vowel space than short vowels, in line with other cross-linguistic accounts of vowel quality differences in long-short vowel pairs (Kristoffersen, 2000, p. 16). In contemporary literature regarding the sound system of Norwegian, the existence of qualitative differences between long and short vowels has

become widely accepted. However, a detailed description of the formant structure of long and short vowels in Norwegian has yet to be created; many popular accounts of Norwegian vowel quantity either do not include a detailed acoustic description or make use of an exceptionally small data set (e.g., Kristoffersen's description being based on his own personal recordings).

Furthermore, mid-vowels /eː/, /oː/, and /øː/ are described as changing from monophthongs in their short form to diphthongs in their long form; the direction of the diphthongization has been described both as centering (i.e., moving toward the center of the vowel space) and opening (i.e., moving toward the edge of the vowel space) (Kvifte & Gude-Husken, 2005). While both long and short vowels in Norwegian can occur in closed syllables, it is worth noting that only long vowels can occur in open syllables.

Table 1.3: Ratio of long to short vowels according to various sources (from Stausland Johnsen, 2019)

Source	**Ratio**
Fintoft (1961)	1.5-2
Vanvik (1972)	2.1-2.7
Payne et al. (2017)	1.4-3.3

In addition to differences in vowel duration and quality, postvocalic consonant duration has been observed to be longer after short vowels than long (Behne et al., 1996). Durational contrasts in Norwegian have been described as more subtle than in languages like Finnish and Hungarian (see Aoyama, 2001, p. 94-96; Ham, 2001, p. 142-150), two

languages where consonant length is not correlated with vowel length and is the main or only carrier of a particular contrast (Payne et al., 2017). The exact ratio of long to short vowels has varied throughout the literature, with scholars claiming a long-to-short ratio of anywhere from 1.0 to 1.8; it is worth noting that the variation in descriptions is smaller than that for vowel durations.

Table 1.4: Ratio of long to short consonants according to various sources (from Stausland Johnsen, 2019)

Source	Ratio
Fintoft (1961)	1.1-1.2
Jensen (1962)	1.2
Vanvik (1972)	1.1-1.4
Payne et al. (2017)	1.0-1.8

Because vowel and consonant length are negatively correlated and one can be predicted from the other, there has been much discussion on whether it is vowel length or consonant length that is phonologically marked; this discussion is not limited to Norwegian, but occurs in closely related languages Swedish and Icelandic, which exhibit "Stress-to-Weight"[2] (Fretheim, 1969; Jahr & Lorentz, 1983). On one side, it is argued that it is vowel duration that is phonologically marked as the durational differences are larger and, therefore, more perceptually salient for listeners (Fintoft, 1961; Behne et al., 1998). On

[2] The so-called "Stress-to-Weight" condition refers to the mutual dependency of syllable weight and stress. This is applicable in quantity discussions in situations where vowel and consonant duration are together said to be dependent on syllable weight.

the other side, some consonants are marked as moraic with vowel length being derived via lengthening under stress in the phonology (Eliasson, 1985; Riad, 1992). Riad (1992) argues and analysis that consonant quantity can, in a straightforward manner, predict the "quantitative complementarity" of segments in stressed syllables, whereas vowel duration fails to do so as long vowels can occur in open syllables. An alternative solution to this debate claims that neither vowel nor consonant length is primary: Kristoffersen (2000, p. 157-158) asserts that vowel and consonant length are both assigned after stress assignment and must both therefore be absent from underlying representations.

Through this dissertation, we will be working under the understanding that vowel quantity is phonologically marked, with differences in consonant length resulting from the quantity of the preceding vowel.

3.3 Perception of Norwegian vowel quantity

3.3.1 Previous work

Perceptual studies on Norwegian about the role of primary and secondary acoustic cues in phonemic quantity during spoken word comprehension are sparse and lack strong claims in their conclusions. Nylund and Behne (1996) examined the salience of vowel quality and duration for Norwegian speakers listening to Norwegian and English vowels. They presented listeners with vowel tokens that were manipulated along both a durational and a spectral dimension and asked listeners to identify the vowel-phoneme they heard. They concluded that duration was mainly used by Norwegians for identifying Norwegian

vowels while quality was inconclusive; the researchers claimed that there was a weak indication that it could be integrated by listeners.

Van Dommelen (1999) looked at temporal factors in the perception of V:C vs. VC: rhymes in Norwegian disyllable words. Van Dommelen used the minimal pair of [ma:tə] (mate) and [mat: ə] (mat, plural adjective) and manipulated them along three dimensions: (1) shortening the long vowel in small steps to create a vowel duration continuum, (2) occlusion duration of the intervocalic consonant in two steps, and (3) original schwa was shortened to create long and short schwa conditions. It is worth noting that the author did not do anything with the quality of the vowel, working under the assumption that spectral differences between long and short /ɑ/ would not influence listener responses. Van Dommelen found that the duration of the vowel had the largest impact on listener quantity perception. Furthermore, a variation in the consonant closure duration appeared to cause a shift in the perceptual long-short vowel boundary, reflecting phonological patterns in Norwegian.

Behne and Nylund (2003) compared Norwegian and English speakers' identification of Norwegian vowels, specifically examining how each listeners group utilized the acoustic cues of vowel duration and quality. Looking at /i, o, a/, they created a 5x5 stimulus matrix design where each vowel was manipulated along both quality and duration dimensions in five steps from an acoustic value canonically associated with a long vowel to a short vowel. The researchers did not include postvocalic consonant closure duration in their investigation and instead normalized this to average duration as produced by the speakers in their study. Listeners heard the manipulated tokens and were instructed to indicate from two real words shown on the screen, which word rhymed

with what they had heard. They found that for /i/ and /o/, vowel quality did not have a significant effect on listener responses for either the Norwegian or American listeners, but vowel duration did for both vowels and groups. Interestingly, the Norwegian listeners used vowel quality in identifying the quantity of /ɑ/, specifically in the form of a large jump from long to short responses between spectral steps 3 and 4. The researchers state that this could be due to the exceptionally large qualitative difference between long and short /ɑ/ as compared to /i/ and /o/ used in this study. In addition to introducing the notion that vowel quality could be used in the perception of quantity in Norwegian, this study also offers the first glimpse at possible vowel-specific perceptual patterns within the language as well.

Taken together, these studies provide evidence that the secondary acoustic cues present in the production of Norwegian vowel quality are salient and informative for listeners: they are useful in quantity perception. However, previous studies have not included a comprehensive subset of the vowel inventory, as in the case of Nylund and Behne (1996) and Behne and Nylund (2003). Considering the emergence of vowel-specific patterning in Behne and Nylund (2003), another look at a larger section of the vowel inventory would seem to be the clear next step. Furthermore, studies such as van Dommelen (1999) have not considered both vowel quality and postvocalic consonant duration together in the same study. Therefore, it is difficult to paint a clear picture of how these two acoustic cues are used adjacently in perception and the possible interactions that may occur.

3.3.2 Preliminary study

In a preliminary study, I looked at the role of vowel duration and spectral quality in the perception of long and short /i, u, a/ in Norwegian. Specifically, I was interested in two main points: (1) is vowel quality used in quantity perception, and (2) is this the same across vowels. Building off previous work that had left agnostic conclusions, I aimed to determine if there was merit in analyzing perceptual patterns for the vowels separately to allow more nuanced patterns to emerge.

Three phonotactically possible CVC non-word minimal pairs differing in quantity containing /i, u, ɑ/ and matched for coda consonant were recorded within the carrier phrase “jeg sa ___ i går” (I said ____ yesterday) by a male native Norwegian speaker. Vowels were spliced out of their original frames and acoustically manipulated along both vowel duration and vowel quality to create a 6x10 stimulus matrix, with six quality steps with F1 and F2 shifting incrementally from the canonically short (step 1) to canonically long (step 6) vowel and 10 duration steps ranging from 70 ms to 160 ms in 10 ms increments. This approach mirrors previous research on cue weighting (e.g., Grenon et al., 2019). The manipulated vowels were put back into their frames, after which the postvocalic consonant closure duration was normalized, where the durational difference between consonant closures after long and short vowels were neutralized via removing or duplicating segments of this part of the recording, to 110 ms, in line with previous work (e.g., Behne et al., 2003).

38 participants (22 female, 15 male, 1 non-binary) were students at the University of Oslo and reported being native speakers of Norwegian. Participants completed a 4AIX paired discrimination task in which each pair contained (1) the stimulus item and (2) either

the unchanged long or short vowel. For example a trial might look like: bi:d/$_{ORIGINAL}$/bVd/$_{MANIPULATED}$ vs. /bɪd/$_{ORIGINAL}$-/bVd/$_{MANIPULATED}$, where the first pair contains the unmodified long vowel and the second pair contains the unmodified short vowel. Listeners were presented with orthographic representations of the word pair and asked to specify what word they heard, categorizing the vowel as long or short.

Listener responses were coded for either a long (=1) or short (=0) vowel categorization and stimuli were coded for both Quality (1-6) and Duration (1-10) steps. The values of the steps were scaled to have endpoints of 0 and 1 in order to made the effect size of the two variables with varying numbers of steps directly comparable. A mixed-effects logistic regression model was run using the *glmer()* function in the *lme4* package in R (Bates et al., 2015). Fixed effects of the model included Quality, Duration, and their interaction and random effects included by-Listener random intercepts and by-Listener random slopes for the main effects and their interaction. Separate mixed effects logistic regression models were run for each vowel phoneme.

Table 1.5: Model output with Coef. (*p*).

	/i/	**/u/**	**/ɑ/**
(Intercept)	**-0.365 (0.022)***	-0.121 (0.659)	**-1.034 (<0.001)*****
Quality	**0.743 (<0.001)*****	**0.594 (<0.001)*****	0.093 (0.431)
Duration	**0.452 (<0.001)*****	**0.912 (<0.001)*****	**0.944 (<0.001)*****
Quality*Duration	**-0.045 (0.031)***	**0.083 (<0.001)*****	**-0.053 (0.003)****
*Syntax: glmer(Response ~ Quality*Duration + (1 + Quality*Duration | Listener)*			

The models for both /i/ and /u/ showed a significant main effect for both Quality and Duration, indicating that listeners used both acoustic dimensions in vowel quantity categorization. However, the /ɑ/ model found only Duration to be a significant predictor of participant responses, suggesting that listeners were not using vowel quality in identifying long and short /ɑ/. The model outputs suggest a phoneme-specific pattern in which listeners utilize differences in vowel quality in the quantity categorization of high vowels but not the low vowel. Furthermore, there is evidence of vowel-specific cue ordering. As we can use the estimated coefficient to approximate the weight of a particular cue, we can see that for /i/, vowel quality is weighted more heavily than vowel duration, whereas the opposite is true for /u/.

These results are interesting for multiple reasons. First, despite previous work claiming that vowel quality was not reliably used, this data suggested that it is indeed used for at least the high vowels /i/ and u/. Furthermore, a phoneme-specific pattern emerged in which both quality and duration were used for high vowels, but not for the one low vowel used. Of course, whether vowel height is a significant factor in what listeners use in quantity perception cannot be definitively determined from this vowel subset, but this warrants further vowel-specific work. Lastly, in addition to phoneme-specific use of acoustic cues, there was also phoneme specific-cue ordering, suggesting that listeners are able to adjust their cue weighting and ordering by-phoneme for whatever reason including cue informativity or based or based on their distributional experience with cues for that particular vowel.

These preliminary results are the jumping off point for Experiment 2 in this dissertation, which will include more vowels and an additional acoustic cue: postvocalic consonant closure duration.

4. THIS DISSERTATION

This dissertation aims to explore the link between phonetic variation and systems of phonological contrast through the lens of spectral and temporal cues in the production and perception of Norwegian vowel quantity. This dissertation does not look at one specific dialect of Norwegian but includes participants from multiple regions across Norway. This will be achieved through a series of experiments that were designed to reveal different aspects of my overall research question; each experiment is a piece of the overall puzzle of the complex nature of Norwegian vowel quantity.

In Chapter 2, Experiment 1 will focus on the acoustic cues that are present in the production of Norwegian vowel quantity. In order to adequately assess the roles of spectral and temporal cues in enhancement and perception, it is important to first establish the cues present in production. As described in 2.1.1, there are a variety of acoustic cues that are known to signal vowel quantity contrasts in different languages and these cues can be on the vowel or adjacent segments. In this chapter, I will be examining the role of four main cues: (1) vowel duration, (2) vowel quality, (3) postvocalic consonant duration, and (4) vowel inherent spectral change in the diphthongization of mid vowels. This chapter will provide a comprehensive account of the acoustic correlates of vowel

quantity in the production of regular speech as produced by 26 native speakers of Norwegian.

In Chapter 3, Experiment 2 will examine the ways in which the same 26 Norwegian speakers enhance vowel quantity through clear speech. In this chapter, a pseudo-interactive task is created to create a scenario where participants encounter an interlocutor who either: (1) correctly understand the intended utterance or (2) misunderstand the intended vowel quantity in the utterance. Participants then either confirm or correct the interlocutor's understanding of their utterance, effectively producing regular or clear speech. How the acoustic cues explored in Chapter 2 are expressed in clear speech as compared to regular speech will be explored and the implications of these enhancements on the phonological representation of Norwegian vowel quantity will be discussed. This study will be the first in-depth description of clear speech in Norwegian and an important investigation into how quantity is enhanced specifically in this language.

In Chapter 4, Experiment 3 will explore how primary and secondary acoustic cues are utilized by listeners. This study will address two shortfalls of previous studies that were noted above: (1) the lack of a more comprehensive vowel subset, and (2) not considering both vowel quality and postvocalic consonant duration together in a single study. Building off the results of my preliminary study looking at the three corner vowels, this study offers an expanded vowel set with the six vowels /i, u, a, o, ø, e/) as well as the inclusion of a third acoustic cue: postvocalic consonant duration. Using a similar stimuli matrix design to the preliminary study, each vowel is manipulated along both duration and quality in five steps to create a 5x5 set. Each set of vowels was then spliced back into the frame with either a long or short postvocalic consonant, created 50 tokens for each vowel

and 300 tokens overall. Participants then heard each stimulus and then was asked to choose from a pair of recordings of real words, which one the stimulus rhymed with, similar to the task used in Behne and Nylund (2003). The way that the acoustic cues are used, interact, and how these strategies might be vowel-specific will be outlined and elucidated.

In Chapter 5, I will discuss the overall findings of Experiments 1, 2, and 3 within both a language-specific descriptive context and how patterns found in Norwegian compare to established cross-linguistic patterns for vowel quantity contrasts outlined by previous research. I will also explore the broader theoretical implications of this research for our understanding of linguistic systems. Specifically, I will address topics such as the role of phonetic variation in system of phonological contrast. Together my dissertation will provide a comprehensive account of the roles of both spectral and temporal cues in this quantitative contrast and describe the link between production and perception. In essence, this is an account of the link between phonetic systems and systems of contrast: how do segmental and suprasegmental phonetic realizations represent and enhance phonological contrasts?

CHAPTER 2 – PRODUCTION OF VOWEL QUANTITY

1. BACKGROUND

Oftentimes, a single acoustic dimension is not sufficient to define a phonological category (Liberman et al., 1967). For example, while VOT is a primary acoustic cue signaling word-final stop voicing, this contrast is often enhanced by a secondary cue of duration on the preceding vowel (Lisker and Abramson, 1964). Therefore, we often see phonological categories marked by multiple, presumably mutually enhancing, secondary cues aimed at making contrasts more salient for listeners. Vowel quantity is when the duration of a vowel encodes lexical distinctions, in that whether a vowel is long or short can change the meaning of a word. Cross-linguistically, vowel quantity is a multi-dimensional and robust phonological contrast with a handful of acoustic correlates that signal it. Common secondary acoustic cues include spectral differences, dynamic f0, and the duration of the following consonant (Maddieson, 1984, p. 129-130; Behne et al., 1996; Pind, 1996).

In order to explore the role of vowel quality, vowel duration, and postvocalic consonant duration in Norwegian phonological quantity, it is important first to establish the acoustic features present in production. In Norwegian, our understanding of vowel quality works has been hampered by two main issues. First, the overall lack of high-quality, acoustic-based phonetic descriptions of the vowel quantity system. The literature is sparse and often uses very small data sets (Behne et al, 1996; Kristoffersen, 2000). For example, in Kristoffersen (2000), the majority of the acoustic vowel analysis is done on recordings made of only the author. Secondly, in the sparse studies available, there

are competing accounts of exactly how Norwegian vowel quantity is signaled acoustically. For example, some more traditional approaches to this issue claim that vowel quantity in Norwegian is signaled solely by temporal cues, with cues such as vowel quality not being affected by quantity (e.g., Behne et al., 1996). Other accounts tell a different story, stating that long and short vowels in Norwegian are in fact realized with different acoustic qualities, with long vowels being more peripheral (e.g., Kristoffersen, 2000). Furthermore, there are competing descriptions of long mid vowels: while they have been stated to diphthongize, some literature describes this diphthongization as centering (Kvifte & Gude-Husken, 2005), while some describes it as moving toward the edge of the vowel space (Kristoffersen, 2000).

Therefore, it is important to continue examining the acoustic correlates of Norwegian vowel quantity to create a cohesive and comprehensive description that can be used as a foundation for future research surrounding the phonetics of Norwegian vowels. From here, we can begin to address larger theoretical questions regarding the multidimensionality and robustness of phonological contrast. In many cases, the primary, obligatory cue for a particular contrast is easily identified, but enhancing secondary cues contribute much to increasing distinctiveness and salience. Examining different types of contrasts in different languages allows us to begin to understand how multidimensionality in contrast can work on a larger scale. Furthermore, we can begin to understand the nuanced role that phonetic variation has within phonological categories and representation.

2. RESEARCH QUESTIONS

Experiment 1 aims to provide an acoustic analysis of the realization of Norwegian vowel quantity in a set of six long-short vowel contrasts: /i/, /u/, /ɑ/, /o/, /e/, /ø/. The goal is to describe the multidimensional acoustic correlates of vowel quantity before moving on to examine how these acoustic cues are enhanced in clear speech (Chapter 3) and used in perception (Chapter 4). Specifically, do Norwegian speakers use multiple cues in production to signal quantity contrasts? And if so, what are these cues? Furthermore, can we predict to see that acoustic cues signaling quantity on the vowel are uniform across the vowel system?

Four acoustic cues are examined: (1) vowel duration, (2) vowel quality, (3) postvocalic consonant duration, and (4) degree and direction of spectral movement in diphthongized long vowels. The specific patterns we predict are outlined below.

As vowel duration is the primary and obligatory cue for vowel quantity, we predict that long vowels have a longer duration than short vowels. Furthermore, we predict that consonants following long vowels are produced shorter than those following short vowels. Given both previous descriptions of vowel quantity and theories such as the Duration Ratio Hypothesis, these predictions are uncontroversial.

While the existence of durational differences in vowels and the postvocalic consonant is widely accepted (e.g., Behne et al., 1996), the role of vowel quality and, also, the diphthongization of long vowels needs to be examined further. To investigate differences in (2) vowel quality between long and short vowels, we can check for significant differences along the F1 and F2 dimensions; if there are indeed spectral

differences, we predict that these will manifest as significantly different values for these formants. Specifically, as previous literature has pointed out that long vowels are often more peripheral than short vowels (Maddieson, 1984 p.129-130; Kristoffersen, 2000), we predict that vowels will be closer to the edge of the vowel space when they are long. This peripheralization would look different for each vowel, given that a peripheral movement is “outward” rather than simply “fronter” or “higher”. Therefore, the differences we predict are: long /i/ has a lower F1 and higher F2, long /u/ has a lower F1 and lower F2, long /ɑ/ has a higher F1 and lower F2, long /e/ has a lower F1 and higher F2, long /ø/ has a lower F1 and higher F2, and long /o/ has a lower F1 and lower F2. These differences are roughly illustrated in Figure 2.1.

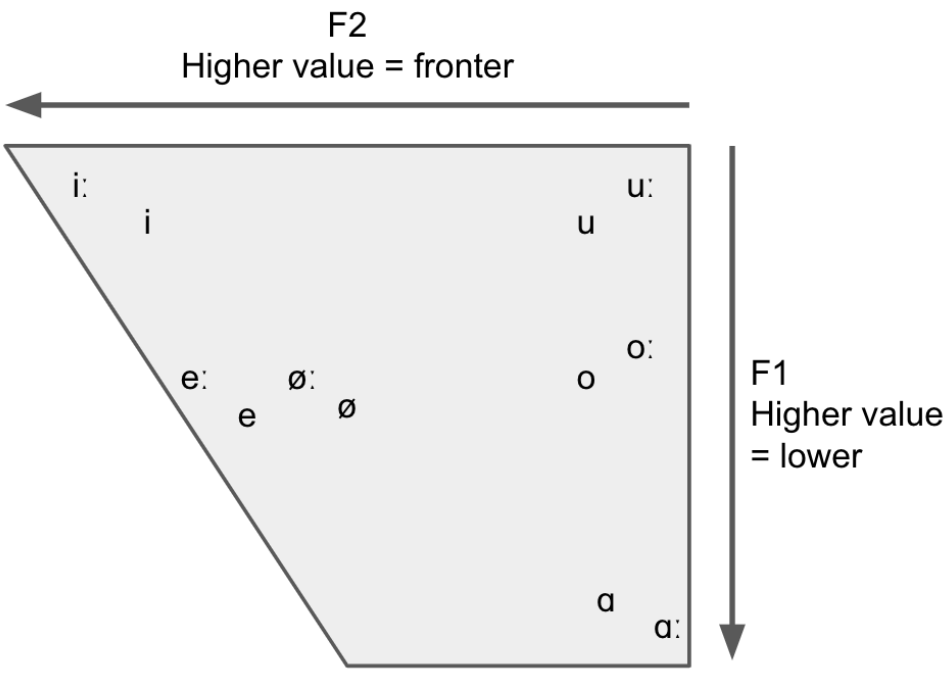

Figure 2.1: Relative predicted positions of long and short vowels within the vowel space.

In addition to placement in the vowel space, we predict that the (4) diphthongization of long mid vowels are used to signal quantity in production. This can be investigated via the amount of spectral movement in long mid vowels. But using the F1 and F2 measurements taken at 20% and 80% of the vowel's duration as points within a F1/F2 two-dimensional space, we can calculate the Euclidean distance to estimate the acoustic distance between the two points, telling us the amount of spectral movement. We predict that long mid vowels will have a larger distance between these two points in the vowel than short mid vowels.

3. METHODS

3.1 Word list

The list consisted of CVC quantitative minimal pairs of real words matched for coda consonant voicing and manner but varying in onset consonant. Table 2.1 provides the word list used in this study. The word list was developed in consultation with a native speaker of Norwegian. In addition to the 12 target words, 24 filler words were also included in the study. Table 2.2 provides these filler words.

Table 2.1. Orthography for target word pairs by-vowel with gloss in parentheses (n=12).

Vowel	Pair
/i/	hvit (white, masc.) - hvitt (white, neut.)
/u/	bok (book) - bukk (ram)
/ɑ/	fat (plate) - fatt[3]
/o/	våt (wet, masc.) - vått (wet, neut.)
/e/	fet (fat, masc.) - fett (fat, neut.)
/ø/	søt (sweet, masc.) - søtt (sweet, neut.)

Table 2.2: Orthography for filler words with gloss in parentheses (n=24).

Mat (food)	Hus (house)	Vin (wine)	Lys (light)
Språk (language)	Bil (car)	Vinn (win)	Øy (island)
Katt (cat)	Venn (friend)	Vind (wind)	Tid (time)
Litt (little)	Melk (millk)	Hunn (female)	Mer (more)
Snill (kind)	Vei (street)	Hund (dog)	Seng (bed)
Sko (shoe)	Fjell (mountain)	Møt (toward)	Takk (thanks)

[3] Typically found as part of a phrase, for example: *ta fatt* (start).

3.2 Participants and procedure

Twenty-six participants (16 female, 9 male, 1 non-binary; average age = 28.3 years) participated in this experiment at the University of Oslo via online recruitment posts. All reported being native speakers of Norwegian and all reported speaking at least one other language other than Norwegian. None of the participants reported having any hearing or speech impairments. The study was approved by the UC Davis Institutional Review Board (IRB protocol #1653463-1) and subjects completed informed consent before participating.

Participants were recorded in a quiet room. Speakers recorded the word list within a frame sentence “Jeg sa ___ i går” (I said ___ yesterday). They were instructed to read the word list as naturally as possible and in their own dialect. Two productions were collected from each participant.

3.3 Acoustic analysis

Each recording was listened to by the researchers in order to ensure that the speaker produced the utterances correctly. Trials with artifacts (e.g., yawning or humming) were exclude; ten trials from two participants were excluded from this experiment. Participant utterances were annotated using a TextGrid in Praat (Boersma & Weenink, 2022) with three tiers (1) production type at the sentence level, (2) word at the word level, and (3) vowel and postvocalic consonant at the segment level.

Two temporal measurements were taken. Vowel duration was measured from the start to the cessation of periodic voicing and clear formant structure. Consonant closure

duration was measured from the cessation of periodic voicing and formant structure in the vowel to either the release burst as seen in the spectrogram or the onset of voicing for the first vowel in the phrase “i går”, after the target word. In addition to temporal measurements, the first two formants were measured from the midpoint of the steady state of each vowel using a script in Praat (Boersma & Weenink, 2022). In order to eliminate differences in the acoustic output based on speaker differences, formant values were normalized using the Nearey1 method (Nearey, 1977), based on log mean normalization with Equation 1. Here, $F^*_{n[v]}$ is the normalized value for $F_{n[V]}$, formant *n* of vowel *V*, and mean(log(F_n) is the log-mean for all F_ns for the speaker in question.

$$F^*_{n[V]} = \text{antilog}\,(\log(F_{n[V]}) - mean(\log(F_n) \qquad (1)$$

These normalized values were then scaled to be more Hertz-like using Equations 2-3 where F^N_i is the normalized value for formant *i* and $F^N_{i\text{MIN}}$ and $F^N_{i\text{MAX}}$ are the minimum and maximum normalized formant values for formant *i* (Thomas & Kendall, 2007).

$$F'_1 = 250 + 500(F^N_1 - F^N_{1MIN})/(F^N_{1MAX} - F^N_{1MIN}) \quad (2)$$

$$F'_2 = 850 + 1400(F^N_2 - F^N_{2MIN})/(F^N_{2MAX} - F^N_{2MIN}) \quad (3)$$

We are also interested in the difference in spectral characteristics between long and short versions of the same vowel phoneme. Therefore, we calculated Euclidean Distance (ED) between the vowels as points on a two-dimensional (F1-F2) plane to approximate the

acoustic distance between them (formula provided in Equation 4), where *a* represents one vowel and *b* represents the other.

$$ED = \sqrt{(F1_a - F1_b)^2 + (F2_a - F2_b)^2} \qquad (4)$$

In addition, measurements of the formants at 20 and 80 percent of the vowel's duration were taken from the mid vowels to examine the degree of vowel inherent spectral movement (VISC) (indicating diphthongization) across long and short vowels.

4. RESULTS

4.1 Vowel duration

Figure 2.2 shows the average duration of long and short vowels by vowel type and Table 2.4 gives the raw duration values for these, as well as the long-to-short ratio for each vowel type. Short vowels had an average duration of 73 ms while long vowels had a duration of 207 ms; the average long-to-short ratio of vowels from this data set was 2.87, within the range of long-to-short vowel durations reported in the literature (1.4 to 3.3).

To assess the effect of both vowel and quantity on the vowel's duration, a mixed effects linear regression was run. Fixed effects included Quantity (Long or Short), as well as Vowel Type (6 levels). Random effects included by-Speaker random intercepts and by-Speaker random slopes for Quantity (see syntax below). Fixed effects were sum coded.

Duration ~ Quantity + Vowel + (1 + Quantity | Speaker) (5)

The output of this model is provided in Table 2.3. As expected, the model showed a significant main effect for quantity, with short vowels being produced reliably shorter than long vowels.

The model did not compute any main effects for vowel type, indicating that there are no significant differences in the average durations of long and short vowels by vowel pair. This is in contrast with findings reported from previous literature, which observed that the low vowel /ɑ/ is produced longer than other vowels in Norwegian.

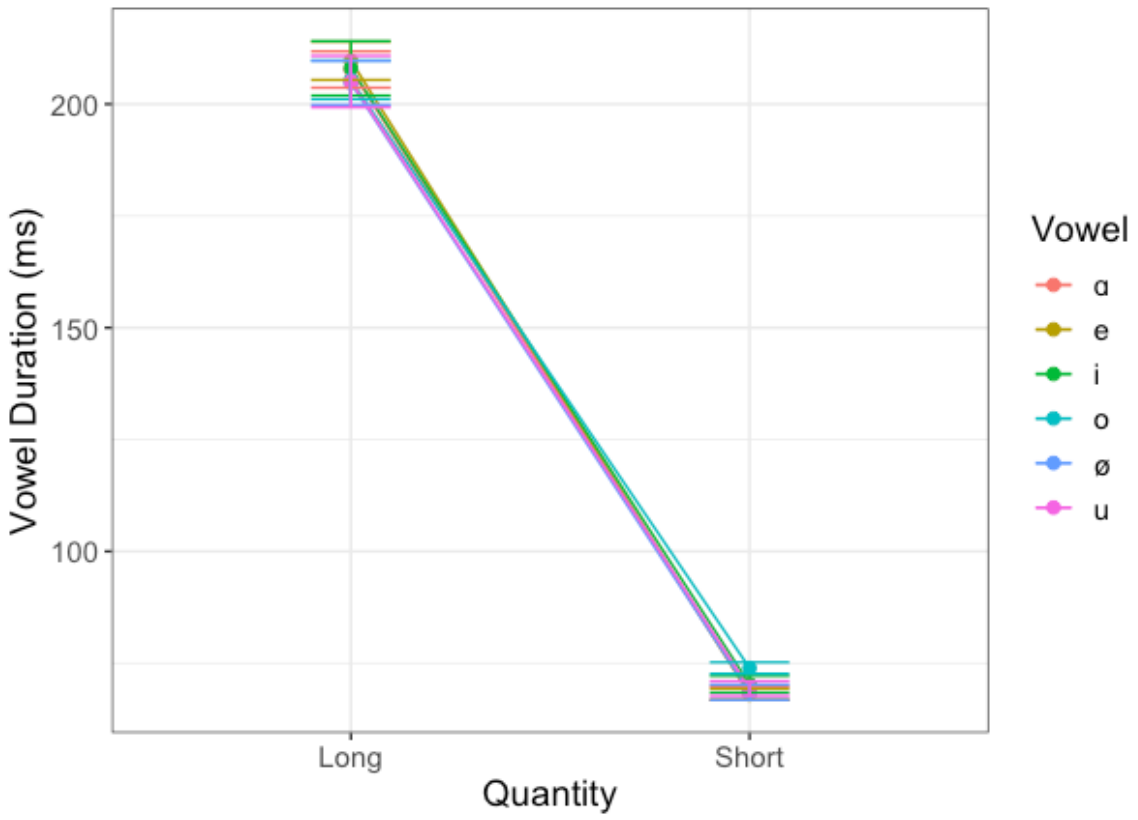

Figure 2.2: Average vowel durations (ms) by-vowel for long and short vowels (a=/ɑ/).

Table 2.3: Model output for vowel duration.

	Est.	Std. Error	df	*t*	*p*
Intercept	**138.517**	**10.051**	**4.000**	**13.781**	**<0.001*****
Vowel (ɑ)	-2.809	5.338	109.000	-0.526	0.599
Vowel (e)	3.003	3.634	109.000	0.563	0.574
Vowel (i)	-0.448	4.389	109.000	-0.084	0.933
Vowel (o)	-1.826	3.123	109.000	-0.342	0.732
Vowel (ø)	-1.793	3.763	109.000	-0.336	0.737
Quantity (Short)	**-65.383**	**2.387**	**109.000**	**27.387**	**<0.001*****

Table 2.4: Average durations for long (A) and short (B) vowels by duration (ms) and the long-to-short ratio (C).

Vowel	**A. Long**	**B. Short**	**C.** $\frac{Long}{Short}$
i	211.020	73.35	2.87
u	211.37	73.76	2.86
ɑ	205.67	72.09	2.85
ø	203.54	70.21	2.89
o	211.02	73.83	2.86
e	210.79	70.48	2.99
Average	**207.21**	**73.31**	**2.87**

4.2 Vowel quality

Figure 2.3 displays the acoustic value of each long and short vowel within the F1-F2 space. Table 2.5 provides the average F1 and F2 values for long and short vowels, by vowel type.

In order to assess whether differences along F1 and F2 were significant, t-tests were conducted for each formant and vowel. For example, to examine which differences are significant for long and short /i/, the difference between long and short F1 was tested as well as the difference between long and short F2. The results of these t-tests can be seen in Table 2.6. For all vowels tested, there was a significant difference along at least one acoustic dimension, indicating that there is a qualitative difference in the production of long and short vowels for all pairs tested. However, which formant—or whether it was one or both that differed—is vowel-specific.

For /i/, there is a significant difference in both the F1 and F2 dimensions; long /i/ has a lower F1 and higher F2, indicating that it is produced higher and fronter in the mouth. For /u/, there is a significant difference only in F1, with long /u/ having a lower F1, indicating that it is articulated higher in the mouth. The lack of significant difference in F2 points to it being articulated with similar backness in long and short forms. The vowel /ɑ/ has a significant difference in F2, indicating that the long vowel was produced higher in the mouth; there is no significant difference in F1 to suggest differences in frontness. For /ø/, there is a significant difference in F1, which was lower in the long vowel, indicating an articulation higher in the mouth. There is no significant difference in F2. The vowel /o/ has a difference along the F1 dimension; long /o/ is produced with a lower F1, pointing to an

articulation higher in the mouth. Lastly, /e/ has a significant difference along both F1 and F2. Long /e/ is produced with a lower F1 and higher F2, indicating a higher and fronter articulation.

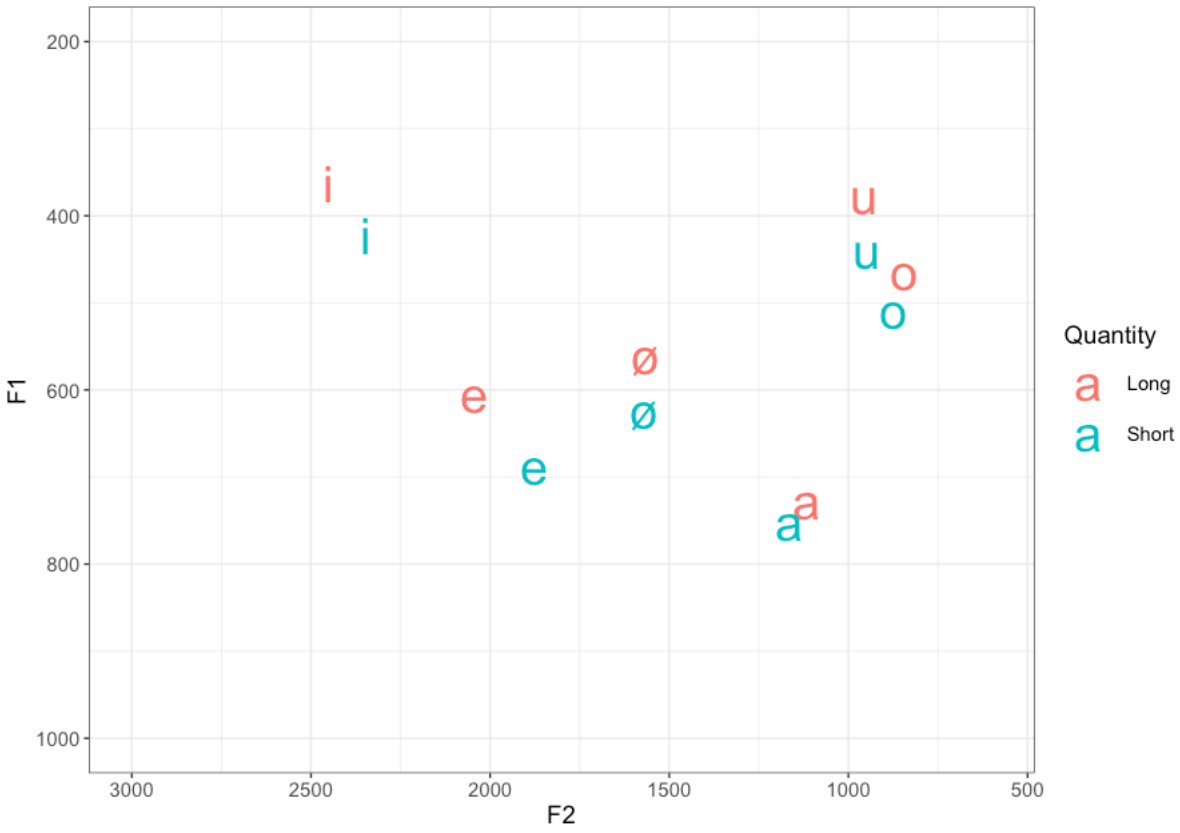

Figure 2.3: Plot of scaled normalized F1 and F2 values for long and short vowels.

Table 2.5: Average formant values in Hz for long and short vowels taken at midpoint.

Vowel	**Long**		**Short**	
	F1	**F2**	**F1**	**F2**
i	362	2451	422	2345
u	376	955	438	946
ɑ	717	1117	752	1165
ø	560	1564	623	1569
o	463	845	508	875
e	604	2044	687	1875

Table 2.6: Results of t-tests and the Euclidean Distance between long and short vowels.

Vowel	**Formant**	***t***	**df**	***p***	**ED (Hz)**
i	**F1**	**-3.238**	**42.821**	**0.002****	122
	F2	**2.092**	**43.965**	**0.042***	
u	**F1**	**-3.161**	**40.971**	**0.003****	63
	F2	-0.119	43.996	0.905	
ɑ	F1	-1.083	42.924	0.2844	55
	F2	**-2.061**	**42.834**	**0.043***	
ø	**F1**	**-3.175**	**43.930**	**0.003****	64
	F2	-0.107	43.651	0.915	
o	**F1**	**-2.481**	**43.078**	**0.017***	54
	F2	-0.712	42.036	0.480	
e	**F1**	**-3.812**	**41.970**	**<0.001*****	188
	F2	**3.637**	**41.833**	**0.001*****	

In addition to the overall placement in the vowel space indicated by the first two formants, the acoustic distance between vowels in each pair was tested. Of the six vowel, four vowels have similar Euclidean distances, within the 54-64 Hz range: /u/ (63 Hz), /ɑ/ (55 Hz), /ø/ (64 Hz), and /o/ (54 Hz). Two vowels have notably larger Euclidean distances between long and short vowels: /i/ (122 Hz) and /e/ (188). To assess the effect of Vowel on the distance between short and long vowels, a mixed effects linear regression was run. Fixed effects included Vowel Type (6 levels). Random effects included by-Speaker random intercepts and by-Speaker random slope. The output of this model is provided in Table 2.7. Fixed effects were sum coded.

ED ~ Vowel + (1 + Vowel | Speaker) (6)

The model shows a significant effect for Vowel, with both /i/ and /e/ having a significantly larger acoustic distance within the long-short pairs than the other vowels. This makes sense, given the patterns in the number of formant dimensions vowel pairs differed on. For the four first vowels with smaller acoustic differences within-pair, there is a significant difference along only one formant. For the two with higher acoustic distances, they differ along two dimensions. Here it is important to note that this data suggests that while there was an acoustic difference found in each of the six long-short pairs in the current study, the extent of that acoustic difference was not uniform.

Table 2.7: Model output for Euclidean distance.

	Est.	Std. Error	df	*t*	*p*
Intercept	**91.100**	**17.350**	**19.84**	**10.428**	**<0.001*** **
Vowel (ɑ)	-54.080	37.862	259.180	-1.428	0.1544
Vowel (e)	**97.392**	**34.341**	**259.180**	**2.932**	**0.012***
Vowel (i)	**31.459**	**45.299**	**259.180**	**2.428**	**0.025***
Vowel (o)	-37.393	36.793	259.180	-1.475	-0.141
Vowel (ø)	-27.452	35.202	259.180	-0.880	0.379

4.3 Postvocalic consonant duration

Figure 2.4 plots the average postvocalic consonant duration after each vowel by quantity; it should be noted that on the x-axis, “long” refers to long vowels, not a long consonant. Table 2.8 has the raw durations in milliseconds for each of these values

As expected, consonants are consistently shorter after long vowels (146.63 ms) than following short vowels (210.55 ms) and this difference is significant [t(5.089) =-3.280, p=0.021). Previous literature reports that consonants following short vowels were 1.0-1.8 times the duration of consonants following long vowels and the present data paints a similar picture: consonants following short vowels are produced 44% longer than those following long vowels. Generally, there is not a difference in the postvocalic consonant duration by nucleus vowel, except for consonants following short /o/, which are slightly longer than consonants following the other vowels.

Additionally, the vowel-to-consonant ratio was measured to examine the proportion of the syllable rhyme that was taken up by each segment type. For words

containing long vowels, the duration of the vowel is 1.43 times that of the consonant whereas in words containing a short vowel, the vowel's duration is 0.35 times that of the following consonant. This is relatively stable across vowels.

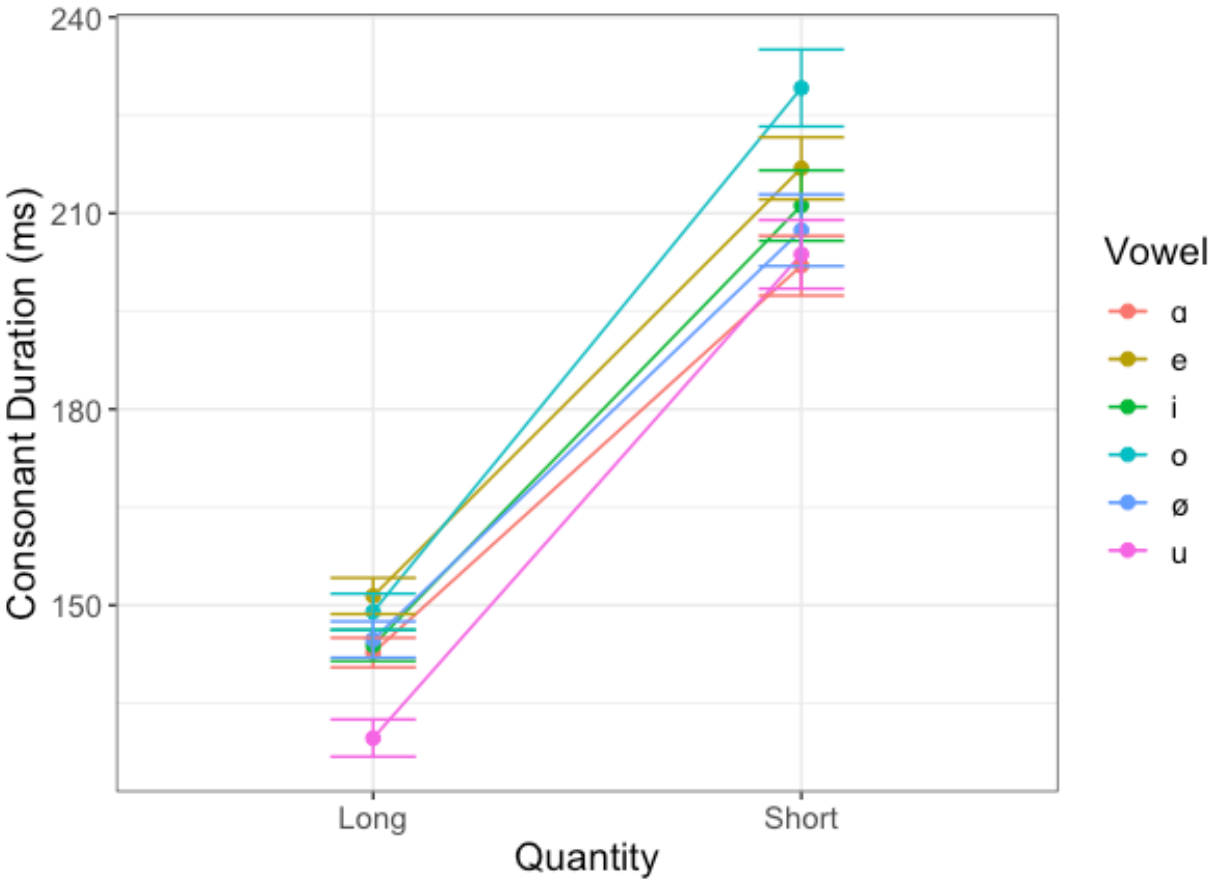

Figure 2.4: Average postvocalic consonant durations after long and short vowels, by vowel type.

Table 2.8: Average durations (ms) for long and short vowels (A), consonants after long and short vowels (B) and the vowel-to-consonant ratio (C).

	A. Vowel	**B. Consonant**	**C.** $\frac{Vowel}{Consonant}$	**D.** $\frac{Long}{Short}$
Long	207.21	146.63	1.43	1.44
Short	73.31	210.55	0.35	

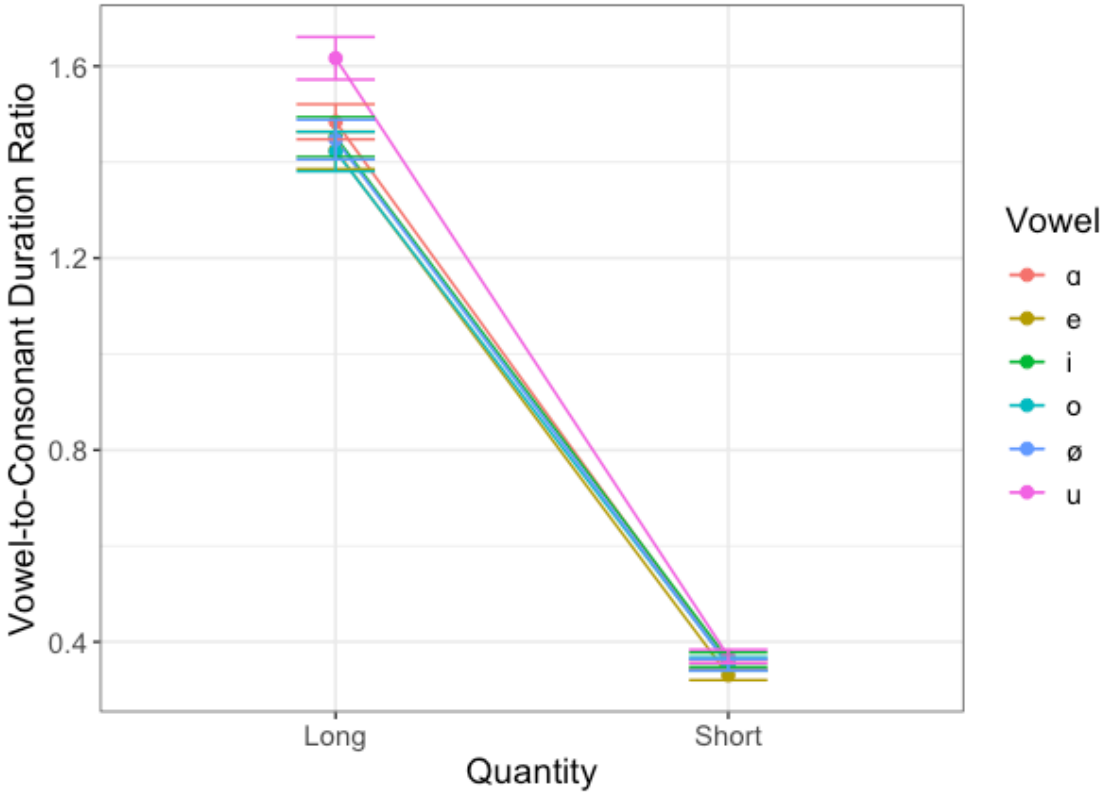

Figure 2.5: Average VC duration ratios by-vowel for long and short vowels.

4.4 Diphthongization (VISC)

For mid vowels, the first and second formants were measured at 20% and 80% of the total vowel duration to assess the degree and direction of diphthongization. Figure 2.6

shows the movement of each long vowel through the vowel space from the start to end of production. Table 2.10 shows the average F1 and F2 values at the start (20%) and end (80%) of each vowel as well as the acoustic distance between these two points.

To assess whether the degree of movement within long mid vowels was greater than within short mid vowels, spectral movement was analyzed with a linear regression model. The fixed effect of the regression was Quantity. Random effects included by-Speaker random intercepts and by-Speaker random slopes for Quantity. Fixed effects were sum coded.

$$ED \sim Quantity + (1 + Quantity \mid Speaker) \quad (8)$$

The output of this model is provided in Table 2.9. Separate models were run for each mid vowel. The decision to run vowel-specific models was in order to look at whether fixed effect (Quantity) was different from zero (if speakers diphthongized) rather than the overall average. As expected, the model for each vowel showed a significant main effect for Quantity for the three mid vowels, indicating that the degree of spectral movement is greater in long mid vowels than in short mid vowels. However, the model did not compute a significant difference for non-mid vowels: long non-mid vowels were not produced with a larger degree of spectral movement than short vowels, indicating that they are not diphthongized.

Of the three mid vowels showing diphthongization, /o/ has the smallest acoustic distance (116 Hz) from the beginning to end of vowel, indicating a lesser degree of diphthongization than the other two vowels. From start to end, the both the vowel's F1

and F2 increase, indicating a movement down and forward, toward the center of the vowel space. /ø/ has the next smallest degree of movement, with a Euclidean Distance of 645 Hz from the start to finish of the vowel. The vowel's F1 increases while its F2 decreases, indicating movement down and backward in the mouth, also moving toward the center of the vowel space. Lastly, /e/ has the largest degree of movement along the vowel, with a distance of 150 Hz from start to finish. Similar to /ø/, the F1 increases while the F2 decreases across the vowel, indicating centralization.

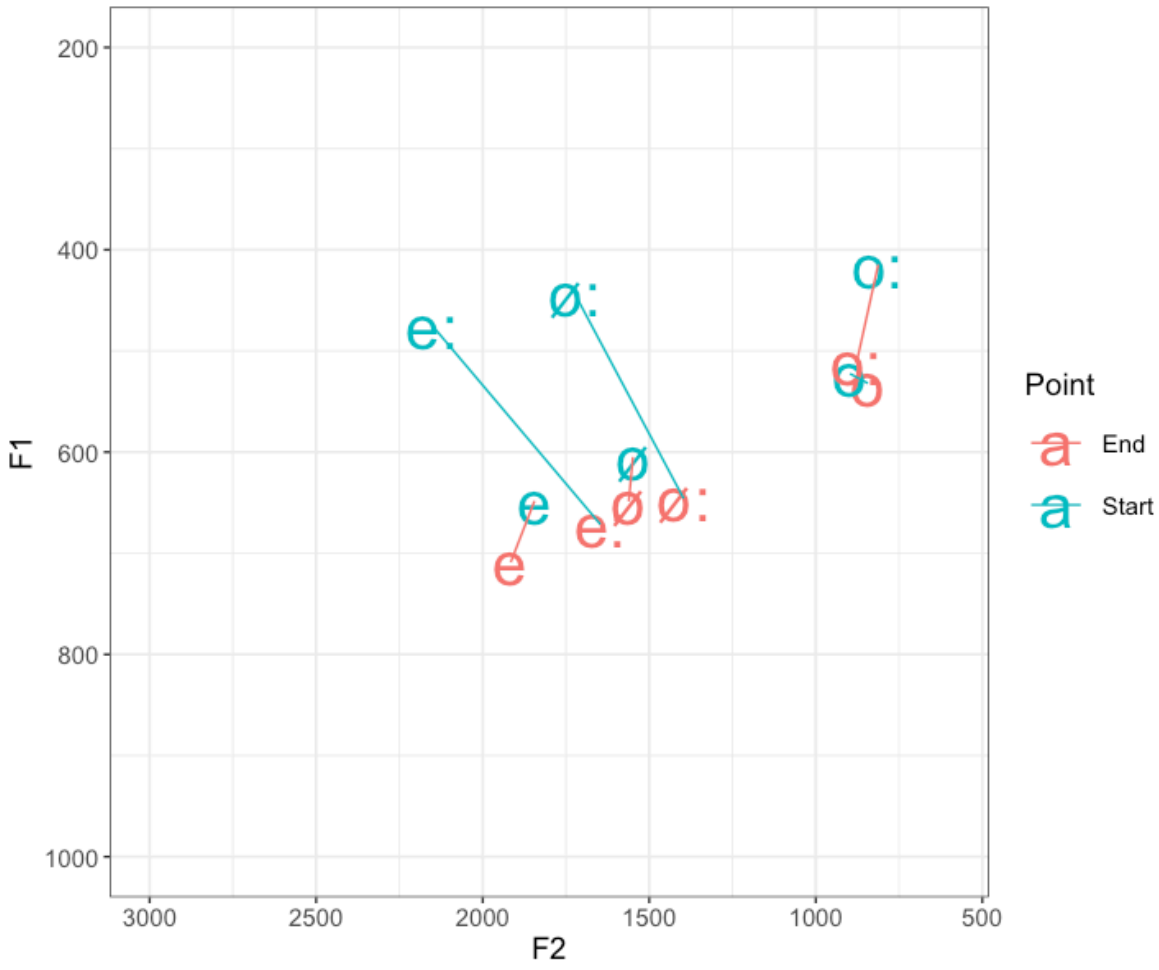

Figure 2.6: Movement of mid vowels through the vowel space.

Table 2.9: Model output for spectral movement.

Vowel		**Est.**	**Std. Error**	**df**	***t***	***p***
i	**Intercept**	**174.430**	**6.454**	**19.173**	**27.030**	**<0.001*** **
	Quantity (Short)	-46.429	29.593	26.393	-1.490	0.093
u	**Intercept**	**118.538**	**9.322**	**20.768**	**12.716**	**<0.001*** **
	Quantity (Short)	-51.331	18.342	25.329	-1.109	0.193
ɑ	**Intercept**	**125.890**	**10.580**	**19.543**	**11.890**	**<0.001*** **
	Quantity (Short)	-75.394	24.301	25.992	-0.356	0.542
e	**Intercept**	**729.340**	**39.300**	**21.590**	**18.560**	**<0.001*** **
	Quantity (Short)	**-621.760**	**41.310**	**29.170**	**-15.050**	**<0.001*** **
o	**Intercept**	**166.36**	**12.65**	**21.860**	**13.152**	**<0.001*** **
	Quantity (Short)	**-73.990**	**20.760**	**22.240**	**-3.564**	**<0.001*** **
ø	**Intercept**	**313.410**	**12.620**	**21.835**	**24.830**	**<0.001*** **
	Quantity (Short)	**225.390**	**21.860**	**26.934**	**-10.310**	**<0.001*** **

Table 2.10: Average formant values for the start and end of long /e, ø, o/ and the Euclidean distance between points.

Point	**Formant**	**/e/**	**/ø/**	**/o/**
Start	F1	475	443	414
	F2	2150	1724	812
End	F1	671	645	511
	F2	1646	1396	877
Euclidean Distance		540	384	116

5. INTERIM DISCUSSION

5.1 Temporal cues

The two temporal cues examined in this acoustic analysis were vowel and postvocalic consonant duration. Given that vowel duration is the primary, obligatory cue to vowel quantity distinctions, it was expected that long vowels would be significantly longer in duration than short vowels. Previous literature provided a range of long-to-short ratios ranging from as small as 1.4 to as long as 3.3 (Fintoft, 1961; Vanvik, 1972; Payne et al., 2017). In the current data set, long vowels are an average of 2.87 times the duration of short vowels, with a range of 2.85 to 2.99 across vowel types. This is within the line of previous research on vowel durations. Another noteworthy point is the lack of vowel-specific differences in average duration observed in the present study. Previous literature, specifically Behne et al. (1996), stated that low vowels were longer on average than high vowels. In their study, they were comparing /ɑ/ to /i/ and /o/. This pattern is not attested in the data here. This is somewhat surprising as intrinsic differences in duration across vowel types are common cross-linguistically, often an articulatory consequence of needing to open the jaw further for low vowels than high vowels (Solé & Ohala, 2010); opening the jaw further simply takes more time. However, we do not see such a trend here. At the present time we cannot provide a conclusive answer as to why, but it is noteworthy nonetheless.

The second temporal cue of interest was postvocalic consonant closure duration. As with many other quantitative languages, Norwegian displays a compensatory trade-off in quantity in that consonants after long vowels are shorter than those after short vowels. The reported durational ratio from previous literature varies, as with vowels, though to a lesser extent. Across various sources, consonants after short vowels are reported as being anywhere from 1.0 to 1.8 times the duration of consonants after long vowels. In the current data set, consonants after short vowels average a duration 1.4 times longer than those after long vowels, comfortably within the ranges provided by previous research. As we know that quantity is often cued by multiple acoustic cues—one of which often being postvocalic consonant duration—this difference in consonant duration after long and short vowels was expected.

In addition to examining the durations of quantitative vowels and their following consonants, it is also worth examining how durations within syllable rhymes are influenced by quantity. To this end, the vowel-to-consonant ratio and overall VC durations were calculated. As immediately evident from the difference in range of durational ratios between vowels and consonants, consonant duration does not vary to nearly the extent that vowel duration does. This is likely due to the fact that contrasts are carried on the vowel and, though postvocalic consonant duration is an enhancing secondary feature, it is not obligatory or as informative in communication. Therefore, we cannot expect to see a truly compensatory lengthening or shortening of consonants within the VC environment leading to symmetrical vowel-to-consonant ratios. In V:C clusters, the vowel comprises, on average, 58.6% of the cluster, while in VC: clusters, the consonant comprises, on average, 74.2% of the cluster. This tells us that while the majority of the VC: cluster's

duration is taken up by the consonant, in the case of long vowels, the consonant isn't shortened to the extent needed to create a similar ratio between V:C and VC: clusters. Of course, this conclusion comes from a relatively limited set of data: monosyllabic words with stop codas. It would of course be of interest in the future to look more closely at how this works in other environments such as multisyllabic words, and words with varying coda consonants.

5.2 Spectral cues

Two main aspects of the spectral quality of the vowels are under investigation in the current study. First, whether there were differences in the spectral characteristics of long-short vowel pairs, and second, if there was diphthongization of long mid vowels as suggested in previous literature.

It has been well established that differences in spectral qualities within long-short vowel pairs is common cross-linguistically within systems that have phonological vowel quantity (Maddieson, 1984, p. 129-130). However, it is not always the case that every vowel pair within a given language has these differences, as evidenced by Maddieson (1984); there are some vowel pairs—or perhaps areas in the vowel space—that show spectral differences in long-short pairs than others. Therefore, it was of particular interest to examine the spectral characteristics of each vowel pair individually.

The first main takeaway from the data presented here is that every vowel pair in the data set has significant acoustic differences along at least one formant, with two vowels having differences in both F1 and F2. The fact that all vowels have a difference

along at least one formant supports the claim that all long vowels are produced with a different quality than short vowels. While this qualitative difference in long-short vowels pairs is attested cross-linguistically, we sometimes see that a qualitative difference exists in some vowel pairs, but not in all. As seen in Maddieson (1984, p. 129-130), there are some areas of the vowel chart that are “hot spots” for this qualitative difference in within the set of languages that have a long-short pair there, some areas have a higher percentage of pairs with qualitative differences. For example, Maddieson (1984, p. 129-130) showed that in high front long-short pairs (40 pairs), 42.5% had spectral differences. On the other hand, areas such as low central had 31 attested long-short pairs, but of these, none had spectral differences. However, in the data presented in this study, all vowels have spectral differences along at least one formant dimension, indicating that spectral differences between long and short vowels is both robust and informative in carrying this phonological contrast in Norwegian.

In addition to exploring if and how the long and short vowels within a pair differed, it was also important to consider the degree of difference, as expressed through the Euclidean Distance between the vowels in the two-dimensional acoustic plane. As demonstrated above, the way in which the vowel pairs differ qualitatively is vowel-specific, the degree to which they differ was as well. Specifically, two vowels stand out in their exceptionally large acoustic difference between the long and short vowel: /i/ and /e/. It is worth noting that these are the only two vowels that have significant acoustic differences along both formant dimensions as well. The case of acoustic differences in /i:/-/ɪ/ contrasts has been the topic of numerous studies looking at the interaction of vowel quality and quantity in production and perception. Researchers such as Kim et al. (2020) often point

to this particular pairing of sounds as being particularly salient in both production and perception. This is the case in both qualitative contrasts, such as tense-lax distinctions where duration is a secondary cue, and in quantitative differences, such as long-short contrasts, where quality is a secondary cue. Some languages even point to quantitative differences in this vowel pair as behaving differently than other vowel pairs. For example, in Czech, listeners will often produce long and short /i/ with a much larger difference in vowel quality and smaller difference in duration than pairs (Podlipský et al., 2009). The exact reason why has yet to be uncovered, but in the context of Norwegian and this dissertation, it will be of interest to see if these differences continue to persist in clear speech enhancement and cue weighting in speech perception.

Lastly, it is noteworthy to revisit descriptions of other languages with vowel quantity produced with spectral differences between long and short vowels. In many languages, long vowels are described as being more peripheral than short vowels (Maddieson, 1984, p. 129-130). The same has also been said about long and short vowels in Norwegian (Kristoffersen, 2000). However, it is unclear if this statement can be broadly applied to the vowels in the current data. It is true that some of the vowel pairs have a long vowel that is clearly more peripheral than the short vowel. For example, long /i/ is produced both fronter and higher in the vowel space and long /u/, while not backer, is still produced higher. Yet, there are vowels in this data set that are either unclear in the peripherality of the long vowel or are very much the opposite in pattern. For mid vowel /ø/, while the long vowel is placed higher in the vowel space, it is not necessarily further front or back and it is ambiguous if this can truly be considered more peripheral. Low vowel /ɑ/ seems to go against this pattern all together with long vowel being produced further from the edge of

the vowel space with a lower F1. Thus, from the current data, it is not clear if the one-size-fits-all “long vowels are more peripheral” statement can truly be applied.

The second point of investigation looked at the way long mid vowels (i.e., /e, o, ø/) are realized in Norwegian. Previous literature stated that mid vowels tend to diphthongize when they are long. However, there were competing accounts as to the degree and direction of this diphthongization, specifically whether the phonetic diphthongs were opening (i.e., moving toward the edge of the vowel space) or closing (i.e., moving toward the center of the vowel space). First, it was established that long vowels are produced with more spectral movement than short vowels; this supports claims that mid vowels tend to diphthongize when they are long, but not when they are short. Therefore, this increase spectral movement is likely helpful in signaling quantity. The data presented in this chapter support claims of a closing diphthongization in that long mid vowels move toward the center of the vowel space. Of the three mid vowels, it is noteworthy that while they all exhibit central-directed movement, the degree to which they diphthongize varies. The two front mid vowels /e/ and /ø/ have a higher degree of diphthongization as evident in both the vowel plot in Figure 2.5 and in the Euclidean distance measurements. This suggests that while it is clear that mid vowels do indeed diphthongize, the degree to which they do so is vowel-specific.

5.3 General remarks

The original goal of this experiment was to establish how the following cues were used in production to signal vowel quantity in Norwegian: (1) vowel duration, (2) vowel quality, (3)

postvocalic consonant duration, and (4) diphthongization of long mid vowels. From the data here, it is evident that vowel duration is the main acoustic correlate and is mutually enhancing with postvocalic consonant duration. In other words, consonants are shorter after long vowels than short vowels. Spectral characteristics of vowels have also been proven to signal vowel quantity. While it is not necessarily the case that long vowels are always more peripheral than short vowels, we can see from the data here that there is indeed a spectral difference in long-short vowel pairs. Lastly, it is the case that long mid vowels tend to diphthongize. While previous literature provided competing accounts about the direction of this diphthongization, the data here support an account of centralizing diphthongization. Therefore, we can conclude that all four acoustic cues examined here are correlated with quantity in production.

CHAPTER 3 – ENHANCEMENT OF VOWEL QUANTITY

1. BACKGROUND

According to Lindblom's (1990) H&H Theory, speech production is adaptive and exists along a continuum between hypo- and hyperspeech, with speakers dynamically adjusting their productions to meet communicative demands. Previous studies have aimed to discover global characteristics of clear speech and have found a number of patterns including: higher amplitude (Picheny et al., 1986), slower speaking rates and increased sentence durations (Picheny et al., 1986; Bradlow, 2002), and greater ranges in f0 (Bond et al., 1989; Summers et al., 1988). There has been further evidence that adjustments in clear speech can be more segment-focused. One well-documented example being the expansion of the vowel space in clear speech (Chen, 1980; Picheny et al., 1986; Moon & Lindblom, 1994; Bradlow, 2002; Krause & Braida, 2004). Temporal cues in languages with long-short vowel distinctions can also be enhanced, as evidenced by changes in durational differences in long-short vowel pairs in Croatian where long vowels were lengthened (Smiljanic & Bradlow, 2008). Overall, cross-linguistic work has found that hyperarticulation may have some global features, but also elicits some local contrast-specific patterns of phonetic modification; this suggests that hyperarticulation is at least partially defined by the enhancement of contrast-specific featural contrasts.

One line of research has examined how speakers adjust their speech in a very specific communicative context: when an interlocutor misunderstands them. This type of clear speech scenario, error resolution, examines how specific and controlled clear

speech can be. Some feature of speech in clarifying misheard speech are deemed global, for example durational increases in vowels and voiceless stops (Ohala, 1994) or increased duration of speech segments (Oviatt et al., 1998). However, there is also evidence that speakers can also produce more “focal” hyperarticulation. For example, speakers have been shown to hyperarticulate VOT in situations where they perceive the possibility for perceptual confusability of a stop’s voicing (Buz et al., 2016). This work suggests that rather than being a stable mode of speaking, hyperarticulation is a more targeted and flexible adaptation.

Examining clear speech and error resolution can begin to answer a number of big-picture questions. Because much of clear speech is produced in the context of a real or perceived communicative difficulty, how speakers adjust their speech for better intelligibility can give us insight into what speakers believe is helpful in understanding their intended linguistic message. For example, the global adjustments outlined above might be seen by speakers as making speech overall easier to follow, parse, and understand. However, the fact that speakers adjust their speech in targeted ways when an interlocutor mishears their intended linguistic message indicates that speakers have systematic ways to adjust their speech that are tailored to the nature of the misunderstanding to attempt to make their speech more easily understood. For example, if an interlocutor misunderstands the intended voicing of a consonants and the speaker enhances the VOT in clarifying, this indicates that the speaker, when taking the perspective of the listener, believes this particular cue to be useful in distinguishing this contrast. Therefore, examining how speakers enhance their speech is useful for understanding what is informative and necessary in signaling a contrast.

The way that the cues associated with Norwegian vowel quantity are enhanced in clear speech has not been explored. Furthermore, the enhancement of vowel quantity in clear speech has been overall understudied. Therefore, the investigation of how vowel quantity in enhanced my Norwegian speakers provides a benefit on two fronts, both in understanding Norwegian specifically, but also in understanding how vowel quantity is enhanced. The findings from this study can be further applied to our understanding of how phonetic features signal phonemic categories and how their adjustment and enhancement can tell us about phonological contrast and the dynamic adaptiveness of speech.

2. RESEARCH QUESTIONS

Experiment 2 investigates the way in which speakers manipulate featural distinctions signaling vowel quantity when trying to clarify misunderstood speech. For example, if a speaker produces the word *våt* with a long /o/, but the interlocutor understands this as *vått* with a short /o/, how will the speaker adjust their speech when correcting? Furthermore, what we find in this experiment can begin to paint a picture about how multidimensional contrasts are enhanced to increase intelligibility.

There are four main acoustic cues that were explored in Experiment 1: (1) vowel duration, (2) vowel quality, (3) postvocalic consonant duration, and (4) long mid vowel diphthongization. Of these four cues, each was shown to be correlated vowel quantity in production. Therefore, in Experiment 2, we ask how these cues are enhanced in clear

speech for error resolution and what this can tell us about the how long and short vowels are mentally represented and what their phonetic targets may be.

If it is the case that these acoustic cues are enhanced in clear speech to make vowel quantity contrasts more salient for listeners, there are a number of patterns we can predict within the data. First, if listeners enhance vowel quality differences, we predict that long and short vowels are produced more differently in clear speech than in regular speech. That is, there might be a larger acoustic difference as told by the Euclidean distance in the vowel plane between them. Second, if speakers enhance postvocalic consonant durations, we predict there will be changes in the relative duration ratio between consonants after long and short vowels, also indicating that they are produced more differently in clear speech. Specifically, because the duration ratio is calculated by dividing the consonant after a short vowel by the consonant after a long vowel, we predict a higher number for this measurement in clear speech than in regular speech. In clear speech, it is common to see segments lengthened overall. However, here we are interested in testing whether the relative duration difference within pair changes, indicating enhancing durational differences between long and short vowels in relation to one another. Lastly, if speakers enhance the diphthongization of mid vowels to mark long vowels, we might expect to see more spectral movement, marked by a larger Euclidean distance between the 20% and 80% points in the vowel's production.

From this data, we can learn about what is seen as useful in making the vowel quantity more salient for listeners who misunderstand the intended message from the speaker. In doing so, we can begin to learn more about what is part of the mental

representation of long and short vowels in Norwegian and how to apply this information to theories of contrast and enhancement on a broader scale.

3. METHODS

3.1 Interlocutor recordings

To elicit more natural speech, a pseudo-interactive task was developed in which participants were told they would be communicating with an interlocutor who would be giving them feedback on the words they were producing. This was chosen over a simple reading task as the goal was to create a more authentic communicative context; this has been shown in previous literature to yield speech that is better understood (Scarborough & Zellou, 2013).

To create the voice of the apparent interlocutor, a wordlist was pre-recorded by a 30-year-old male native speaker of Norwegian from Larvik. Words were recorded inside of the carrier phrases of “Var det __ du sa?” (*Was it __ you said?*) and “Ok, du sa __” (*Ok, you said __*). Sentences were altered to include a 100 ms pause before and after the target word and the intensity was normalized to 60 dB for all utterances.

The list consisted of CVC quantitative minimal pairs of real words matched for coda consonant voicing and manner but varying in onset consonant. Table 3.1 provides the word list used in this study. The word list was developed in consultation with a native speaker of Norwegian. In addition to the 12 target words, 24 filler words were also included in the study. Table 3.2 provides these filler words.

Table 3.1: Orthography for target word pairs by-vowel with gloss in parentheses (n=12).

Vowel	**Pair**
/i/	hvit (white, masc.) - hvitt (white, neut.)
/u/	bok (book) - bukk (ram)
/ɑ/	fat (plate) - fatt[4]
/o/	våt (wet, masc.) - vått (wet, neut.)
/e/	fet (fat, masc.) - fett (fat, neut.)
/ø/	søt (sweet, masc.) - søtt (sweet, neut.)

Table 3.2.: Orthography for filler words with gloss in parentheses (n=24).

Mat (food)	Hus (house)	Vin (wine)	Lys (light)
Språk (language)	Bil (car)	Vinn (win)	Øy (island)
Katt (cat)	Venn (friend)	Vind (wind)	Tid (time)
Litt (little)	Melk (millk)	Hunn (female)	Mer (more)
Snill (kind)	Vei (street)	Hund (dog)	Seng (bed)
Sko (shoe)	Fjell (mountain)	Møt (toward)	Takk (thanks)

[4] Typically found as part of a phrase

3.2 Participants and procedure

Twenty-six participants (16 female, 9 male, 1 non-binary; average age = 28.3 years) participated in this experiment at the University of Oslo via online recruitment posts. All reported being native speakers of Norwegian and all reported speaking at least one other language other than Norwegian. None of the participants reported having any hearing or speech impairments. The study was approved by the UC Davis Institutional Review Board (IRB protocol #1653463-1) and subjects completed informed consent before participating.

Participants completed a speech production elicitation task. On a given trial, participants completed the experiment in four parts, schematized in Figure 3.1. First, participants produced the target word within the frame sentence “Jeg sa ___ i går” (I said ___ yesterday). After this initial production, participants heard one of two response conditions from the pre-recorded utterances by the interlocutor: (1) the correct word response or (2) an incorrect word with the incorrect vowel quantity; these response words were played within the frame “Var det ___ du sa?” (Was it ___ you said?). Depending on which condition the particular trial was, participants responded with either: (1) “Ja, jeg sa ___ i går” (Yes, I said ___ yesterday) or (2) “Nei, jeg sa ___ i går” (No, I said ___ yesterday). The last part of the trail consisted of the interlocutor confirming the word the participant produced by saying “OK, du sa ___” (OK, you said ___) with the correct word.

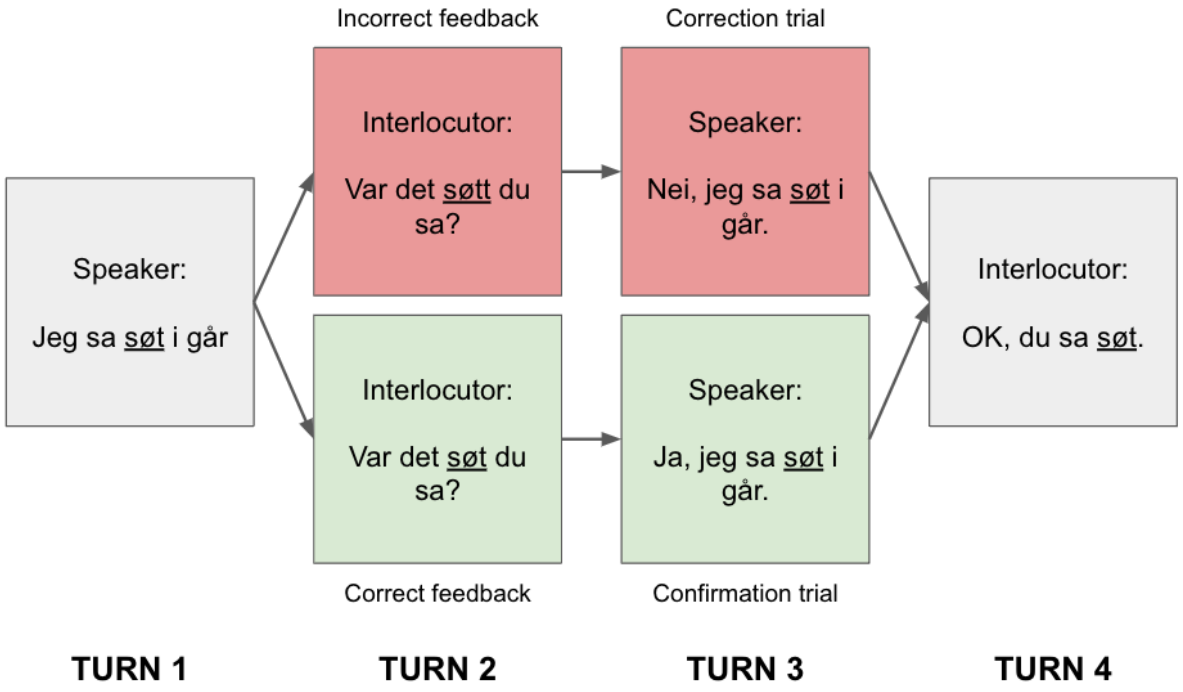

Figure 3.1: Schematic of two possible trials with the word *søt*; a correction trial flow on top and a confirmation trial flow on bottom.

3.3 Acoustic analysis

The same acoustic measures in Experiment 1 were taken in Experiment 2: (1) vowel duration, (2) consonant closure duration, (3) F1 and F2 at vowel midpoint, and (4) F1 and F2 at 20% and 80% of the vowel duration for mid vowels. The same vowel normalization technique described in Experiment 1 was also used for Experiment 2. For the spectral measures, the Euclidean distance between the midpoint of long and short vowels was calculated using Equation 2, where *a* represents one vowel and *b* represents the other. The same equation was also used to calculate the distance between the 20% and 80% points in mid vowels for assessing spectral movement.

$$ED = \sqrt{(F1_a - F1_b)^2 + (F2_a - F2_b)^2} \quad (2)$$

For both duration types, a duration ratio was calculated. For vowel duration ratios, the duration of the long vowel was divided by the duration of the short vowel. For the postvocalic consonant closure duration, the duration of the consonant after a short vowel was divided by the duration of the duration of the consonant after a long vowel.

3.4 Statistical analysis

Each acoustic parameter (vowel duration ratio, Euclidean Distance, consonant closure duration ratio, VISC) was analyzed using a separate linear mixed effects regression model with the *lme4* package in R (Bates et al., 2015). The fixed effect of Production Type was also included (3 levels): (1) *original*: the first production of the sentence prior to feedback, (2) *correction*: repetition of the sentence following incorrect feedback, and (3) *confirmation*: repetition of the sentence following correct feedback. For all models, the initial utterances were used as the reference level. Random effect structure included Speaker random slopes and intercepts for Type (*lmer* syntax provided in Equation 3).

Feature ~ Type + (1 + Type | Speaker) (3)

For features on the vowel (i.e., vowel duration, vowel quality, and diphthongization), a separate model was run for each vowel; a single model was run for postvocalic consonant duration.

4. RESULTS

4.1 Vowel duration

Figure 3.2 shows the average vowel duration for long and short vowels for each trial type and Table 3.3 shows the average long-to-short duration ratio for each vowel and each trial type. Table 3.4 has the statistical output for the linear regression model run for each vowel. The decision to run vowel-specific models was in order to look at whether fixed effect (Type) was different from zero (e.g., if speakers enhanced) rather than the overall average.

For this analysis, the duration ratio of long vowels divided by short vowels is calculated. Because it is long divided by short, a larger ratio indicates a larger relative difference in durations between long and short vowels and a smaller ratio indicates a smaller relative difference.

The models for each vowel showed a similar output and will therefore be discussed here together rather than individually. For each model, there was a significant main effect of Type for both confirmation and correction trials. For the confirmation trials, the negative estimated coefficient tells us that the duration ratio is smaller and, thus, long and short vowels are produced with a duration that is less different in these trials. In contrast, the

positive estimated coefficient for correction trials shows that in clear speech produced in these trials, the relative duration ratio is larger than in the initial utterances, denoting that listeners are enhancing differences in vowel duration in clear speech.

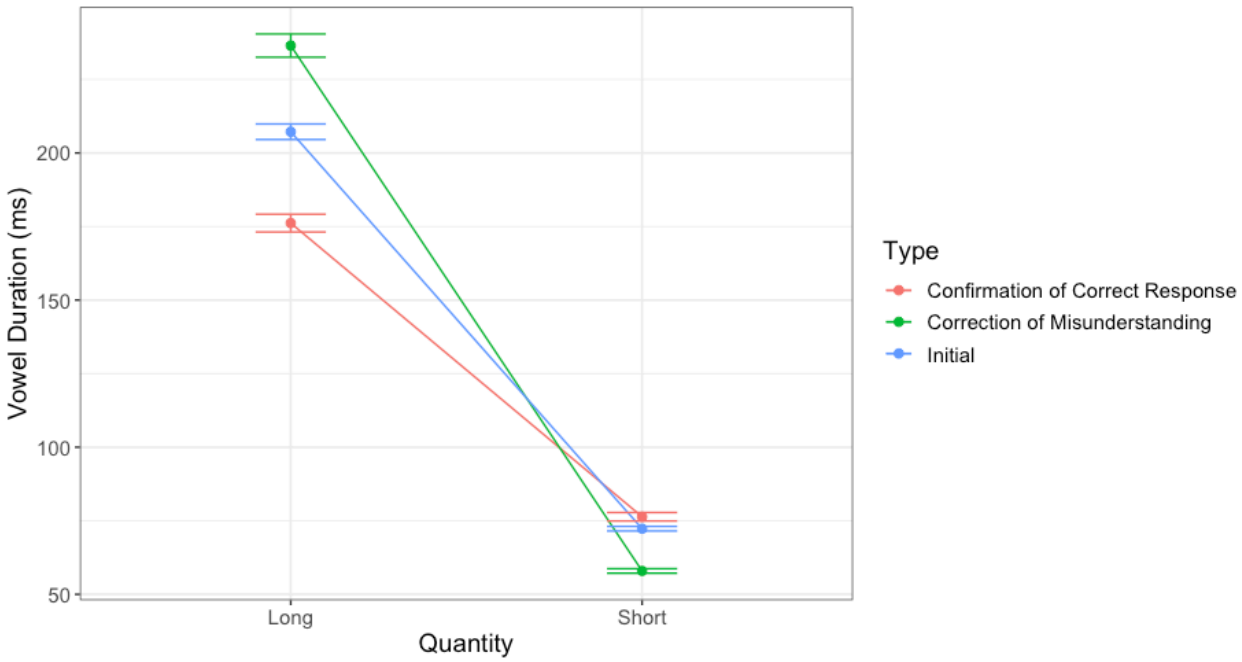

Figure 3.2: Average vowel duration (ms) across trial types.

Table 3.3: Average long-to-short duration ratio by trial type.

Vowel	Initial	Correction	Confirmation
i	2.87	4.31	2.37
u	2.86	3.72	2.65
ɑ	2.85	4.31	2.57
ø	2.89	4.51	2.30
o	2.86	4.06	2.08
e	2.99	4.41	2.42
Average	2.87	4.22	2.40

Table 3.4: Model outputs for vowel duration.

Vowel		Est.	Std. Error	df	*t*	*p*
i	**Intercept**	**2.982**	**0.217**	**23.008**	**13.708**	**<0.001*** **
	Type (Confirm)	**-0.601**	**0.134**	**26.728**	**-4.573**	**<0.001*** **
	Type (Correct)	**1.331**	**0.129**	**44.281**	**10.246**	**<0.001*** **
u	**Intercept**	**2.962**	**0.143**	**23.603**	**20.663**	**<0.001*** **
	Type (Confirm)	**-0.305**	**0.118**	**47.715**	**-2.591**	**0.013***
	Type (Correct)	**0.758**	**0.163**	**25.548**	**4.629**	**<0.001*** **
ɑ	**Intercept**	**2.941**	**0.132**	**22.792**	**22.022**	**<0.001*** **
	Type (Confirm)	**-0.365**	**0.110**	**38.085**	**-3.311**	**0.002****
	Type (Correct)	**1.366**	**0.167**	**23.489**	**8.205**	**<0.001*** **
ø	**Intercept**	**2.993**	**0.153**	**23.018**	**19.509**	**<0.001*** **
	Type (Confirm)	**-0.689**	**0.183**	**29.814**	**-3.761**	**<0.001*** **
	Type (Correct)	**1.514**	**0.228**	**23.045**	**6.616**	**<0.001*** **
o	**Intercept**	**2.741**	**0.077**	**24.311**	**35.618**	**<0.001*** **
	Type (Confirm)	**-0.665**	**0.084**	**30.969**	**-7.890**	**<0.001*** **
	Type (Correct)	**1.322**	**0.165**	**23.560**	**8.028**	**<0.001*** **
e	**Intercept**	**2.992**	**0.117**	**23.166**	**25.558**	**<0.001*** **
	Type (Confirm)	**-0.569**	**0.085**	**25.930**	**-6.636**	**<0.001*** **
	Type (Correct)	**1.418**	**0.209**	**23.063**	**6.760**	**<0.001*** **

4.2 Vowel quality

Figure 3.3 shows the location of long and short vowels within the vowel space for each trial type. Table 3.5 shows the output for the models run on each vowel. The decision to

run vowel-specific models was in order to look at whether fixed effect (Type) was different from zero (e.g., if speakers enhanced) rather than the overall average. From the models, we can see some vowel-specific patterns in enhancement of quantitative differences, which are discussed below.

In the /i/ model, there was not a significant main effect for Type for confirmation trials, indicating that speakers produce long and short /i/ with similar acoustic distance in both initial and confirmation trials. However, there was a significant main effect for Type for correction trials. Here, listeners produce long and short /i/ with a greater acoustic difference between them, indicating they are enhancing the qualitative difference in clear speech. There was a similar pattern in the /u/ model. There was not a significant main effect of Type for confirmation trials, but there was for correction trials. This suggests that, like /i/, speakers do not necessarily reduce the acoustic distance between long and short /u/ in the reduced speech in confirmation trials but do in the clear speech produced in correction trials.

In the /ɑ/ model, there was not a significant effect of Type for confirmation or correction trials. This suggests that speakers do not produce long and short /ɑ/ more or less spectrally distinctly for either confirmation or correction trials. Thus, we can conclude that speakers are not enhancing spectral difference in the long-short /ɑ/ vowel pair in clear speech.

In the /ø/ model, there was a significant main effect of Type for both confirmation and correction trials. In confirmation trials, speakers produce long and short /ø/ with a larger acoustic distance in the vowel space than in initial trials. This is somewhat surprising as we might expect more hypoarticulation in confirmation trials, which may lead

us to expect a smaller acoustic difference rather than a larger one. However, for correction trials, speaker do produce long and short /ø/ further apart in the vowel space, indicating that speaker do indeed enhance qualitative difference for this pair in clear speech. The output of the /e/ model is similar in the fact that there was a significant main effect for Type for both confirmation and correction trials. The positive estimated coefficients for both these trial types suggests that speakers are producing long and short /e/ more distinctly in the second utterances in both confirmation and correction trials than in the initial utterances. From this, we can conclude that speakers are enhancing qualitative differences for this vowel.

The /o/ model had a significant main effect for Type for both confirmation and correction trials. However, unlike the /ø/ model, the negative estimated coefficients in this model indicated that listeners are producing long and short /o/ with a smaller acoustic distance between them for both trial types. While this is not necessarily surprising for confirmation trials, where we might expect more hypoarticulation and the degradation of some acoustic cues, this was not expected for correction trials. We can therefore conclude that listeners do not enhance spectral differences for /o/ in clear speech.

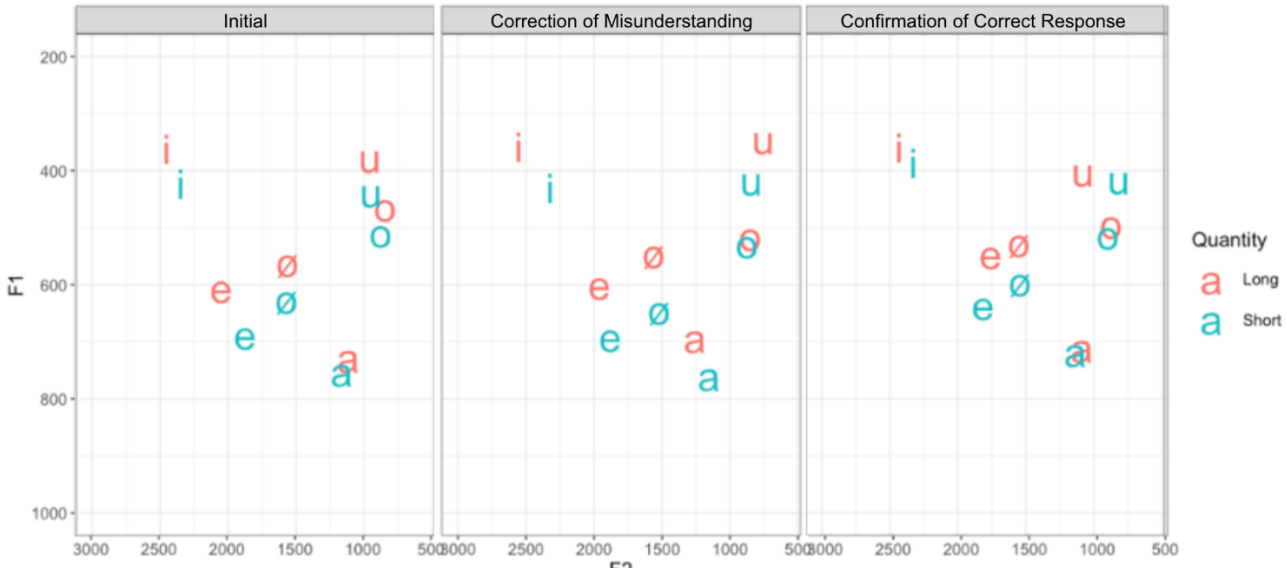

Figure 3.3: Location of long and short vowels within the vowel space by trial type: confirmation (left), correction (middle), or initial (left). (note: a = /ɑ/)

Table 3.5: Model outputs for Euclidean distance.

Vowel		Est.	Std. Error	df	*t*	*p*
i	**Intercept**	**145.000**	**10.226**	**68.047**	**14.180**	**<0.001*** **
	Type (Confirm)	6.961	18.825	42.909	0.370	0.713
	Type (Correct)	**99.496**	**18.455**	**44.538**	**5.391**	**<0.001*** **
u	**Intercept**	**416.680**	**95.230**	**26.470**	**4.376**	**<0.001*** **
	Type (Confirm)	10.400	135.750	51.000	0.077	0.939
	Type (Correct)	**-258.84**	**143.11**	**43.850**	**-1.809**	**0.047***
ɑ	**Intercept**	**127.420**	**11.110**	**48.600**	**11.468**	**<0.001*** **
	Type (Confirm)	-10.620	19.020	50.960	-0.558	0.579
	Type (Correct)	105.310	58.420	21.950	1.803	0.085.
ø	**Intercept**	**94.500**	**8.741**	**24.464**	**10.811**	**<0.001*** **
	Type (Confirm)	**33.260**	**15.416**	**26.704**	**2.158**	**0.040***
	Type (Correct)	**61.101**	**22.406**	**22.180**	**2.727**	**0.012***
o	**Intercept**	**90.044**	**6.032**	**68.751**	**14.928**	**<0.001*** **
	Type (Confirm)	**-27.089**	**10.435**	**69.510**	**-2.569**	**0.012***
	Type (Correct)	**-55.169**	**23.193**	**21.901**	**-2.379**	**0.0265***
e	**Intercept**	**201.660**	**19.130**	**22.570**	**10.543**	**<0.001*** **
	Type (Confirm)	**179.920**	**85.970**	**21.980**	**2.093**	**0.048***
	Type (Correct)	**42.860**	**18.870**	**46.130**	**2.271**	**0.028***

In addition to testing whether the F1/F2 acoustic distance between long and short vowels was enhanced in clear speech, it was also investigated how the vowels changed between initial and correction utterances to induce these larger Euclidean distances. To assess this, t-tests were done for each vowel to assess whether changes in F1 or F2 between

initial and correction utterances were significant and the positivity or negativity of the t-value was used to assess the direction of this change. The t-values and p-values of these tests can be seen in Table 3.6.

For /i/, there was no significant difference in the F1 of the long vowel between initial and correction utterances, yet there was a difference in the F2 values between these two speaking styles. With a higher F2 in correction utterances, this indicates that speakers produce long /i/ more front, or peripherally, in clear speech than regular speech. For the short vowel, there was a significant difference in F2, with a lower F2 value, indicating an articulation that is more back in clear speech than in regular speech. This suggests that in clear speech, speakers are producing short /i/ more centrally than in regular speech.

For /u/, there was a significant difference in both F1 and F2 for the long vowel: a lower F1 and lower F2 indicate that long /u/ is produced higher and backer, or more peripherally, in clear speech than in regular speech. There was no significant difference in either F1 or F2 for short /u/ between the two speaking styles, indicating short /u/ is produced with approximately the same quality in these two speaking styles.

For /ø/, there was not a significant difference in F1 for long vowels between initial and correction utterances, but there was for F2; the lower F2 value indicated that long /ø/ is produced higher in the vowel space by speakers in clear speech than in regular speech. For short /ø/, there was a similar pattern, where there was not a difference in F1 between speech styles, but there was for F2. In this case, a higher F2 value indicated that speakers produce short /ø/ lower in the vowel space, or more centrally, in clear speech than in regular speech.

Lastly, for /e/, there was a significant difference in the F2 values for long /e/ between initial and correction utterances, with a higher F2 value. This indicates that speakers produce long /e/ as more front, or peripherally, in the vowel space in clear speech. No significant difference in F1 tells us that there is not a difference in the vowel's height between these speech types. Furthermore, a lack of significant difference in either F1 or F2 between short /e/ in initial and correction utterances indicates that listeners produce short /u/ in these speech styles with approximately the same value.

Table 3.6: T-test output for differences in F1 and F2 between initial and correction utterances by-vowel.

Vowel	Quantity	F1			F2		
		t	**df**	***p***	***t***	**df**	***p***
i	Long	0.249	33.294	0.804	**2.077**	**33.552**	**0.042***
	Short	-0.377	32.401	0.707	**-2.098**	**32.188**	**0.038***
u	Long	**-1.833**	**32.610**	**0.049***	**-2.338**	**34.104**	**0.022***
	Short	0.903	33.842	0.371	1.384	33.098	0.171
ø	Long	1.0816	31.924	0.284	**-1.932**	**32.790**	**0.044***
	Short	-0.736	33.472	0.467	**2.011**	**32.682**	**0.039***
e	Long	0.177	32.349	0.860	**2.577**	**33.548**	**0.012***
	Short	-1.192	32.949	0.848	0.016	32.582	0.986

Overall, we can see some patterns on the vowel emerge. First, vowel duration is enhanced in clear speech compared to regular speech. This can be seen in the increased long-to-short vowel durations (see Table 3.5), where a larger ratio indicates a larger

relative difference between the long and short vowels. One particularly interesting aspect of this change is that while previous literature has stated that segments tend to be universally lengthened in clear speech, short vowels in the data here are actually shortened instead. This is evidence to support that vowel duration is controlled by the speaker and vowel duration is enhanced here.

The second cue to consider here is vowel quality; this was evaluated via the Euclidean distance between long and short vowels. While vowel duration was enhanced across the vowel subset in a very uniform manner, the enhancement of qualitative differences was not as simple. Here, we begin to see more vowel-specific patterns emerge. Four of the six vowels did show enhancement of qualitative differences via larger Euclidean distances, but two vowels did not. Long and short /ɑ/ did not have any significant difference between them, indicating that speakers do not enhance their qualitative difference, even though it does exist (as seen in Experiment 1). Long and short /o/ are a peculiar case where there is a significant difference in the Euclidean distance between vowels, but the difference is actually smaller rather than larger. This indicates that qualitative differences, although present in regular speech, are not enhanced here. Yet, there was not a clear pattern of how this difference was achieved, whether it meant systematically peripheralizing long vowels, centralizing short vowels, or both. From vowel-to-vowel, there were slightly different patterns in which vowel carried a significant qualitative change along the F1/F2 dimensions. Nonetheless, this data suggests more segment- and even phoneme-targeted enhancement strategies by speakers and sheds light on the complex nature of phonetic feature enhancement for improved intelligibility.

4.3 Postvocalic consonant duration

Figure 3.4 shows the average postvocalic consonant durations for each type of trial type. Here, one should note that “long” does not refer to a long consonant, but rather the consonant after the long vowel. Table 3.7 provides the average postvocalic consonant durations by-vowel and by-type as well as the average across vowels. Table 3.8 shows the statistical output of the model used to assess the significance of durational differences across trial types.

For this analysis, the duration ratio of consonants after short vowels divided by consonants after long vowels is calculated. Because we expect that consonants after long vowels are shorter than consonants after short vowels, the ratio should be above 1.0. Furthermore, because we would expect that enhancement of postvocalic consonant ratio would entail longer consonants after short vowels and shorter consonants after long vowels, a larger relative difference in the consonant durations after each vowel quantity would lead to a larger ratio. On the other hand, a lack of enhancement and underrealization of durational differences in the consonants after each vowel quantity would lead to a smaller relative difference and a smaller ratio.

Rather than running separate models on each vowel as was done for vowel duration and vowel quality, there was one model run to assess the overall changes in consonant duration regardless of the preceding vowel. The model showed a significant main effect for Type for both Confirmation ($p<0.001$) and Correction ($p<0.001$) trials, suggesting that there is a significant difference in the consonant duration ratio between the initial utterances and the correction and confirmation utterances. We can use the

estimated coefficients to tell us both the direction and magnitude of the change. For the confirmation trials, the negative coefficient indicates that compared to initial utterances, the consonant duration ratio is smaller when participants confirmed a correct understanding from the interlocutor. As stated above, the smaller coefficient indicates a smaller difference in the relative durations of consonants after long and short vowels. In other words, the duration of consonants after long and short vowels is less different in these conditions, suggesting a hypoarticulation in this regard. On the other hand, for correction trials, the positive coefficient indicates a larger ratio in correction utterances compared to initial utterances. This larger coefficient points to a larger relative difference, in turn meaning consonants after long and short vowels are being produced more differently than in initial utterances. This suggests that the postvocalic consonant duration is indeed being enhanced in clear speech to help make the contrast being long and short vowels more salient in cases of error resolution.

The size of the relative coefficients can give us information about the magnitude of difference in the ratio. For example, are changes in how different the duration of consonants after long and short vowels are larger in one trial type versus the other? For confirmation trials, the absolute value of the estimated coefficient (-0.088) is smaller than for correction trials (0.134), suggesting that speakers change how differently they produce consonants after long vs. short vowels more in correction trials, where they are hyperarticulating, than in confirmation trials, where they are likely hypoarticulating.

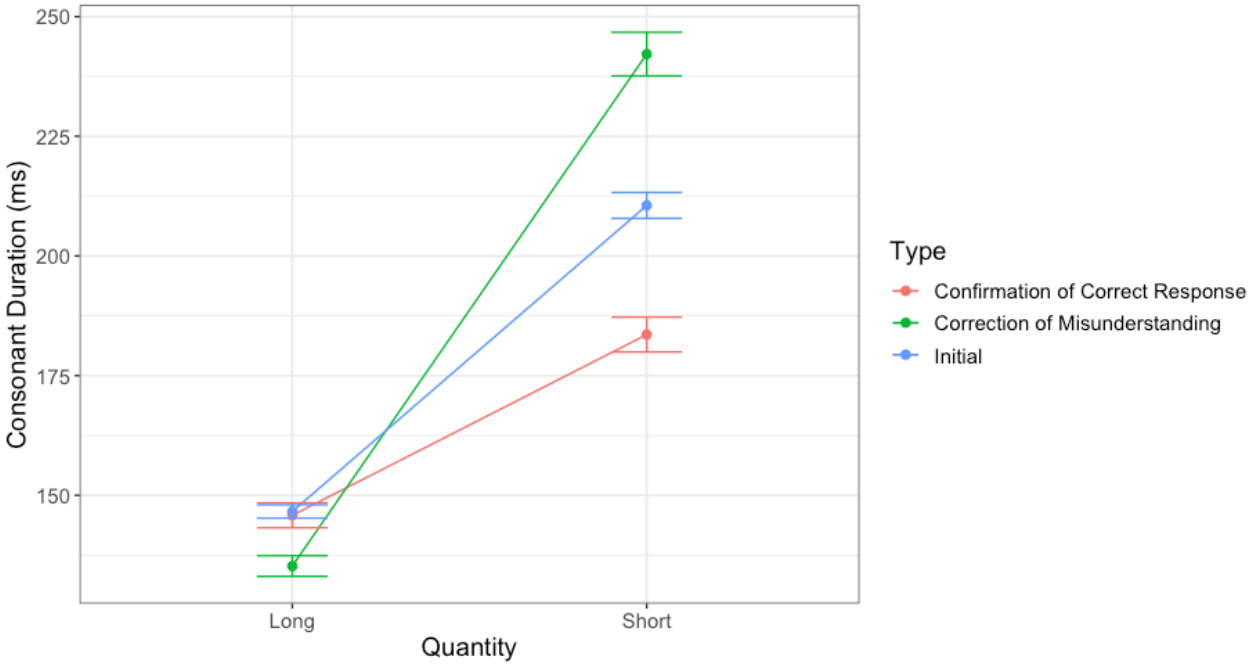

Figure 3.4: Average postvocalic consonant duration (ms) after long (left) and short (right) vowels across trial types.

Table 3.7: Average durations (ms) for long and short vowels (A), consonants after long and short vowels (B) and the vowel-to-consonant ratio (C).

Type	**Vowel Quan.**	**A. Vowel**	**B. Cons.**	**C.** $\frac{Vowel}{Consonant}$	**D.** $\frac{Long}{Short}$
Initial	Long	207.21	146.63	1.43	1.44
	Short	73.31	210.55	0.35	
Confirmation	Long	176.17	145.81	1.24	1.30
	Short	76.36	183.56	0.44	
Correction	Long	236.49	135.22	1.79	1.78
	Short	57.95	242.161	0.25	

Table 3.8: Model output for postvocalic consonant.

	Est.	Std. Error	df	*t*	*p*
Intercept	**0.717**	**0.018**	**23.213**	**39.310**	**<0.001*** **
Type (Confirm)	**-0.088**	**0.013**	**523.653**	**-6.808**	**<0.001*** **
Type (Correct)	**0.134**	**0.012**	**529.996**	**11.546**	**<0.001*** **

Additionally, the vowel-to-consonant ratio was examined to see if the proportion of the syllable rhyme that was taken up by each vowel type was changed in clear speech compared to regular speech. In the initial utterances, the duration of the vowel was 1.43 times that of the consonant in V:C clusters, but this increases to 1.79 in clear speech. This indicates that in clear speech, the vowel is longer compared to the following vowel in V:C segments, suggesting that speakers are enhancing the relative length of the vowel compared to the consonant. In VC: clusters containing a short vowel, the opposite is true. In initial utterances, the length of the vowel was 0.35 the duration of the consonant, but this decreased to 0.25 in correction trials. The smaller ratio indicates that in VC: clusters with a short vowel, the vowel is shorter in relation to the consonant in clear speech, suggesting that speakers are enhancing the length of the consonant compared to the vowel in these rhymes.

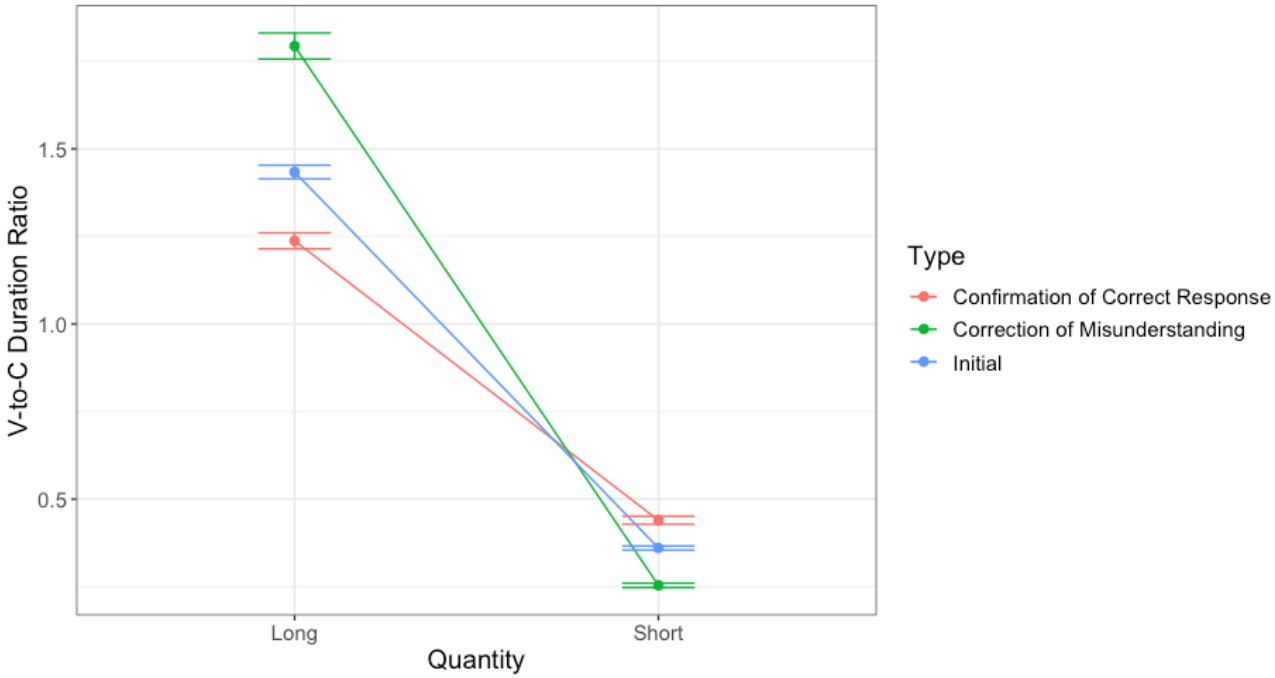

Figure 3.5: Average VC duration ratios by-vowel for long and short vowels by trial type.

4.4 Diphthongization (VISC)

Figure 3.6 demonstrates movement of the long and short mid vowels through the vowel space by trial type. To assess diphthongization of long mid vowels, the Euclidean distance between the 20% and 80% points of the vowel's duration was calculated as a measure of the degree of diphthongization and used in the linear regression model. As with the previous models assessing cues on the vowel, a separate model was run on each of the mid vowels. The decision to run vowel-specific models was in order to look at whether fixed effects (Type and Quantity) were different from zero (if speakers enhanced at all) rather than the overall average. Unlike the above measures, here we are looking at the direct degree of spectral movement rather than a ratio of values relating to long and short

vowels. Therefore, Quantity (long vs. short) is included as a fixed effect in this regression model as well as the interaction between Quantity and Type.

In the /e/ model, a significant main effect for Type and Quantity was found. Overall, the amount of spectral movement within the vowel is lower in confirmation trials and higher in correction trials, indicating that for both long and short vowels, vowels are produced with more spectral movement in clear speech. Furthermore, the significant effect of quantity tells us that across trial and speech types, speakers produce long vowels with more spectral movement than short vowels. The model revealed a significant interaction between Type and Quantity for confirmation trials: in confirmation trials, the difference in spectral movement between long and short vowels is smaller than in initial utterances. In contrast, the model did not compute a significant interaction between Type and Quantity for correction trials: speakers do not produce long and short vowels with a larger difference in spectral movement in correction trials, suggesting that diphthongization of /e:/ is not enhanced in clear speech.

The /o/ model showed a significant main effect of Quantity: across trial types, short vowels were produced with less spectral movement than long vowels. However, there was not a main effect of Type, indicating that there is no overall difference across vowel quantities in the amount of spectral movement based on trial type. Like the model for /e/, there was a significant interaction between Type and Quantity for confirmation trials, indicating that speakers produce long and short /o/ with more similar degrees of spectral movement in these environments. However, the lack of a significant interaction between Type and Quantity for correction trials indicates that, like with the /e/ model, speakers do not enhance the diphthongization of /o:/ in correction trials and clear speech.

The /ø/ model revealed a significant main effect for Type and Quantity. For Type, speakers produce both long and short vowels with an overall smaller degree of spectral movement in confirmation trials and larger degree of spectral movement in correction trials. Speakers also produce a larger degree of spectral movement in long vowels than short, as made evident in the significant main effect of Quantity. The interaction between Type and Quantity was significant for confirmation trials, indicating that speakers produce long and short vowels with more similar degrees of movement in confirmation trials, but the lack of significant interaction for correction trials indicates this is not the case there.

Together, we see a similar pattern for each mid vowel: speakers do not seem to enhance degree of spectral movement, or diphthongization of long mid vowels, in clear speech. Recall in 4.2 where the acoustic difference between long and short vowels at midpoint was discussed.

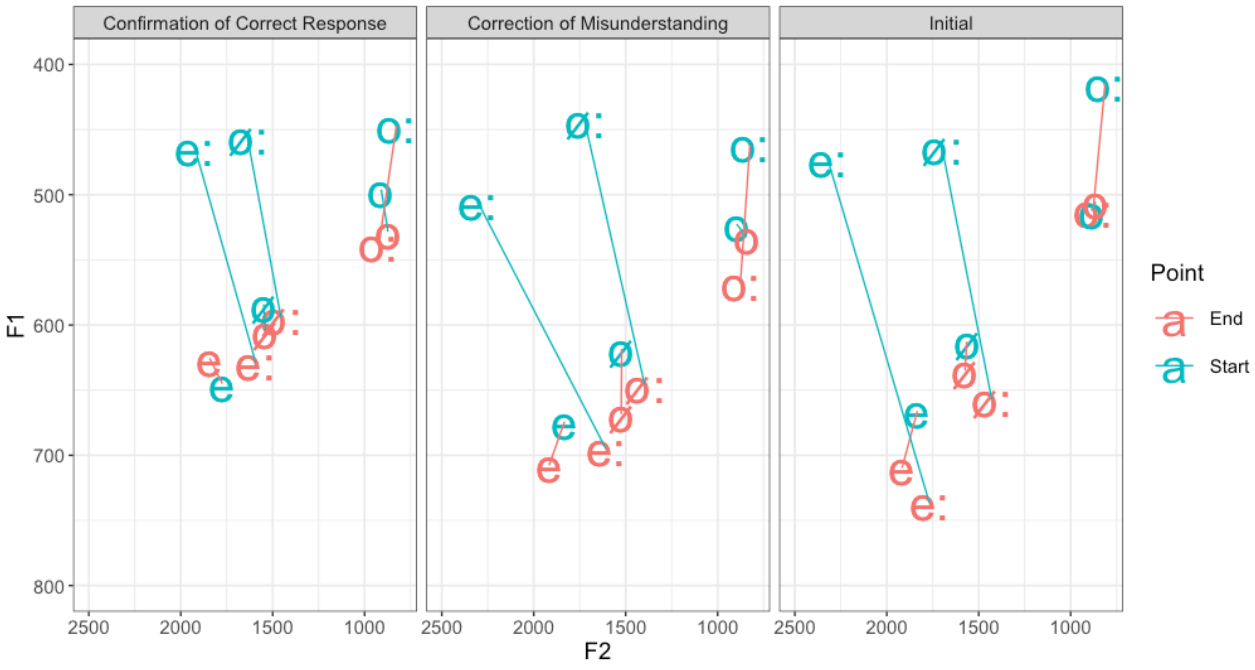

Figure 3.6: Movement of mid-vowels through vowel space by trial type.

Table 3.9: Model outputs for spectral movement.

Vowel		Est.	Std. Error	df	*t*	*p*
e	**Intercept**	**631.140**	**28.640**	**35.79**	**22.039**	**<0.001*** **
	Type (Confirm)	**-223.830**	**44.510**	**163.710**	**-5.029**	**<0.001*** **
	Type (Correct)	**98.150**	**44.510**	**163.710**	**2.205**	**0.029***
	Quantity (Short)	**-512.940**	**36.340**	**55.080**	**-14.115**	**<0.001*** **
	Type (Con) * Quant. (Short)	**-206.540**	**62.940**	**163.710**	**-3.281**	**0.001****
	Type (Cor) * Quant. (Short)	116.190	62.920	163.710	-1.846	0.067.
o	**Intercept**	**166.220**	**12.520**	**27.520**	**13.276**	**<0.001*** **
	Type (Confirm)	-25.320	18.200	163.870	-1.391	0.166
	Type (Correct)	-10.770	18.200	163.870	-0.592	0.555
	Quantity (Short)	**-88.100**	**14.860**	**32.340**	**-5.928**	**<0.001*** **
	Type (Con) * Quant. (Short)	-20.920	25.740	163.870	-0.813	0.417
	Type (Cor) * Quant. (Short)	24.880	25.740	163.870	0.967	0.335
ø	**Intercept**	**360.433**	**9.517**	**50.220**	**37.871**	**<0.001*** **
	Type (Confirm)	**-115.324**	**16.484**	**163.180**	**-6.996**	**<0.001*** **
	Type (Correct)	**32.813**	**16.848**	**163.180**	**1.991**	**0.048***
	Quantity (Short)	**-285.924**	**13.459**	**56.640**	**-21.243**	**<0.001*** **
	Type (Con) * Quant. (Short)	**-128.842**	**23.312**	**163.180**	**-5.527**	**<0.001*** **
	Type (Cor) * Quant. (Short)	-24.747	23.423	163.180	-1.062	0.290

5. INTERIM DISCUSSION

5.1 Temporal cues

In this experiment, the two temporal cues that were examined were vowel and postvocalic consonant duration. Because vowel duration is the primary cue signaling vowel quantity, it was expected that long vowels would be produced with a longer duration than short vowels. This was the case in Experiment 1, where we see long vowels produced with an average duration 2.87 times as long as short vowels. Given the important nature of this cue in vowel quantity production, we expected that it would be enhanced in the clear speech, specifically in the condition where an apparent interlocutor misunderstands the intended vowel quantity. This is indeed the case. The ratio between long and short vowels increases from 2.87 in initial utterances to 4.22 in correction trials. Because this ratio was obtained by dividing the long vowel's duration over the short vowel's duration, the larger ratio confirms that there was a larger relative difference in duration between long and short vowels in clear speech from correction trials. Furthermore, the degree of enhancement of durational differences between quantities is relatively stable across vowels, with long and short /u/ being produced with a slightly smaller increase in durational difference, going from a ratio of 2.86 in the initial utterance to 3.72 in clear speech from correction trials. This finding is interesting when we consider the fact that in clear speech, segments are typically universally lengthened yet in this case, short vowels in clear speech were produced shorter than in regular speech, as highlighted in Figure

3.2. This indicates an intentional, controlled alteration of vowel length, confirming the important role of vowel duration differences in signaling duration.

A difference between initial utterances and confirmation trials, where we might expect to see more hypoarticulation, is also seen. For all vowels, the long-to-short duration ratio decreases from initial utterances to confirmation utterances, indicating that in the speech produced in these trials, long and short vowels are produced with a smaller relative durational difference. This is in line with many of the acoustic characteristics of hypospeech outlined by Lindblom's (1990) H&H Theory.

The second temporal cue of interest was postvocalic consonant closure duration. As with many other quantitative languages, Norwegian displays a compensatory trade-off in quantity where consonants after short vowels are produced longer than those after long vowels. This trend was established in Experiment 1, where consonants after short vowels were produced 1.44 times longer than those produced after long vowels. In confirmation trials, we can see that this ratio drops to 1.30, indicating that the consonants after long and short vowels are being produced with a relatively smaller difference. Critically, this ratio increases in correction trials to 1.78, meaning that speakers are producing consonants after long and short vowels with a larger relative difference in clear speech, indicating that this acoustic cue is enhanced here. Given the informativity of postvocalic consonant duration as demonstrated in Experiment 1, this enhancement of the relative durations was expected. However, as evident in Figure 3.3, the difference in postvocalic consonant duration was carried mostly by consonants following short vowels. The lack of magnitude of change in consonant following long vowels suggests that short

vowels could be more specified for postvocalic consonant length than long vowels. This point would be interesting to address further in future studies.

In addition to examining the durations of vowels and the following consonants, it is worth examining how durations within syllable rhymes are influenced by speech type and if the relative amount of a VC cluster taken up by the vowel or consonant is enhanced in clear speech. First, recall the patterns seen in the initial utterances as described in Experiment 1: in V:C clusters, the vowel comprised 58.6% of the cluster while in VC: clusters, the consonant comprised 74.2% of the clusters. If the relative durations of vowels and the following consonants is used by speakers to make quantity contrasts more salient, we could expect to see a change in the relative ratio of these durations. In clear speech elicited in correction trials, the proportion of V:C clusters that the vowel comprised increases from 58.6% to 63.6% and the proportion of VC: clusters that the consonant comprised increases from 74.2% to 80.9%. What this tells us is that in clear speech, long vowels are longer and short vowels are shorter in direct relation to the duration of the following consonant. This contrasts what we see in the confirmation trials where the proportion of V:C clusters that vowels comprise decreases to 54.8% and the proportion of VC: clusters that consonants comprise decreases to 70.6%. In these trials, long vowels are shorter in relation to the following consonant and short vowels are longer; this suggests that the mutually enhancing nature is diminished in these trials while it is enhanced in clear speech from correction trials.

5.2 Spectral cues

Two main aspects of spectral quality are under investigation in the current study. First, whether or not differences in spectral characteristic of long-short vowel pairs outlined in Experiment 1 are enhanced in clear speech, and second, whether diphthongization via spectral movement in long vowels is enhanced too.

Previous literature has established that differences in spectral quality within long-short vowel pairs is common cross-linguistically within systems of phonemic quantity contrast (Maddieson, 1984, p. 129-130). Experiment 1 demonstrated that this is also the case in Norwegian: every vowel pair is produced with qualitative differences along at least one formant dimension. In four of the six vowels in this experiment, speakers produce long-short pairs with a larger acoustic distance between them in clear speech produced in correction trials. This increased acoustic distance here tells us that when an apparent interlocutor misunderstands the intended quantity of a vowel as produced by a speaker, speakers tend to enhance the spectral difference between long and short vowels for /i, u, ø, e/. It is worth noting that the enhancement of spectral differences in long-short vowel pairs was not uniform across the vowels. For both /ɑ/ and /o/ (discussed further below), speakers do not enhance the acoustic distance between the long and short vowel in clear speech, despite doing it for the other four vowels in the subset. The particular case of spectral differences not being enhanced for /ɑ/ might be explained by general cross-linguistic trends in spectral differences for long-short vowel pairs. As outlined in Maddieson (1984, p. 129-130), long-short pairs in some areas of the vowel space tend to be more prone to having spectral differences than others. For example, 42.5% of high

front long-short pairs and 27% of high back long-short pairs showed spectral differences while 0% of the low back pairs examined in Maddieson's (1984, p. 129-130) did. Given that using spectral differences in distinguishing between long and short low vowels is already rare cross-linguistically, it may be the case that while it is produced in Norwegian, it is not yet a part of the phonetic target and, thus, not prone to being enhanced. The same explanation cannot be applied to /o/; in Maddieson's (1984, p. 129-130) work, 50% of mid back long-short pairs were produced with spectral differences.

The second point of investigation looked at the way diphthongization found on long mid vowels was influenced by speech type. Previous literature stated that mid vowels tend to diphthongize when they are long and this claim was supported by the data presented in Experiment 1. Specifically, long vowels were shown to diphthongize with a centralizing direction (i.e., moving toward the center of the vowel space). Critically, the amount of spectral movement between long and short mid vowels was shown to be significantly different, suggesting that this diphthongization is used to signal long mid vowels in production. In the current study, whether this diphthongization is enhanced in clear speech, specifically via differences in the spectral movement in long and short vowels. Of the three models run, one for each vowel, no model had a significant interaction between Type and Quantity for correction trials, suggesting that the difference in spectral movement between long and short mid vowels is not increased or decreased in clear speech. From this, we can make the conclusion that diphthongization is not enhanced in clear speech.

The changes in temporal cues in consonants and vowels are easily explained by Lindblom's (1990) H&H Theory where listeners increase discriminability by exaggerating

the phonetic differences in a given sound contrast, yet the lack of spectral manipulation for a number of the vowels is unexpected under this account. The vowel /o/ presents us with an especially interesting case in terms of clear speech enhancement. In Experiment 1, long and short /o/ were shown to have both an overall difference in spectral quality and a difference in degree of spectral movement. However, neither of these cues were enhanced in clear speech; in fact, the overall acoustic difference between long and short /o/ was smaller in correction trials than in initial utterances, implying a degradation of this cue rather than enhancement. In the case that the overall spectral difference is not enhanced in clear speech, one might expect that the degree of spectral movement might be enhanced, but this was also not the case. From this, it would seem that for long and short /o/, the contrast is enhanced mainly via temporal cues rather than spectral, similar to /ɑ/.

5.3 General remarks

The original goal of this experiment was to establish ow the following cues are enhanced in clear speech to make the contrast between long and short vowels more salient: (1) vowel duration, (2) vowel quality, (3) postvocalic consonant duration, and (4) diphthongization of long mid vowels. From the data here, it is evident that while all four cues were shown in Experiment 1 to signal vowel quality in production, their enhancement in clear speech is more nuanced. Both the temporal cues are reliably enhanced in clear speech, with increasing long-to-short ratios indicating larger relative differences between long and short vowels and consonants after long and short vowels. Furthermore, the

relative duration of vowels and consonants within rhymes is enhanced, with long vowels comprising more of V:C clusters and consonants comprising more of VC: clusters in clear speech. For vowel quality, we can see that spectral differences are enhanced for some vowels, but not for all. Specifically, we do not see a difference in the acoustic distance between long and short /ɑ/ in clear speech and even a reduction in acoustic distance between long and short /o/. Furthermore, the degree of diphthongization on long mid vowels is not enhanced in clear speech for any vowels. Therefore, we can conclude that temporal cues are more uniformly and reliably enhanced than spectral cues, even though spectral cues are enhanced in a subset of the vowels.

CHAPTER 4 – PERCEPTION OF VOWEL QUANTITY

1. BACKGROUND

From previous work on speech production, it has been demonstrated that a single acoustic dimension is rarely sufficient to define phonological category membership. For example, up to sixteen acoustic dimensions covary with the stop voicing distinction in English (Lisker, 1986). However, speech production is only one half of communication: speech is produced to be perceived. Thus, examining how listeners rely on multidimensional acoustic cues to determine phonemic category membership is helpful to fully understand how phonological contrasts function in a langue (Casserly & Pisoni, 2010). We know that some cues in production are more strongly correlated with category membership than others and, as a consequence, listeners rely on some cues more than others (Abramson & Lisker, 1985; Holt & Lotto, 2006). In one frequently cited example, the English sounds /b/ and /p/ differ systematically in the voice onset time (VOT), but also, though less reliably, in other dimensions such as f0 and stop closure duration. As a result, listener categorization of /b/ and /p/ is mainly determined by VOT, with the value of f0 having a weaker, albeit still present, effect (Abramson & Lisker, 1985). This phenomenon–the ability of listeners to weight information across different acoustic dimensions–has been referred to as perceptual cue weighting (Holt & Lotto, 2006; Francis et al., 2008).

One important question within research on cue weighting has been whether we see a parity between speech production and perception, as in whether the cues that are

strongly correlated with a contrast as produced by speakers are, in turn, important for listeners when perceiving it. How listeners weight cues in perception, even for the same type of contrast, has been shown to vary based on language, context, and individuals (Holt & Lotto, 2006; Clayards, 2018; Schertz et al., 2020). For example, Schertz et al. (2020) examined the perception of stop voicing contrast in Spanish and English, testing monolingual and bilingual speakers' use of four acoustic dimensions: (1) VOT, (2) f0, (3), first formant (F1), and (4) stop closure duration. They found that monolingual English speakers relied more on F1 and less on closure duration than monolingual Spanish speakers, indicating language-specificity in cue use. The language-specific nature of cue weighting in perception suggests that cue use must be to at least some extent based on the distributional properties of cues in production. Neary (1997) demonstrated a way of analyzing speech perception as simple pattern recognition by systematically analyzing acoustic cues of vocoid duration and voice bar duration in /hVC/ stimuli and finding that participants responses mapped with the distributional properties of the input they received.

An account of production cues leading directly to predictable cue weighting in perception would assume a close parity in the production and perception systems and a static relationship between acoustic cues and their respective contrast. What happens, though, when there is a misalignment between cues in production and perception either between or within speakers? A difference between how strongly a cue is correlated with a contrast in production and how heavily it is weighted in perception has been argued by some point to an ongoing cue shift within the language, where cues and their relative correlation with a particular contrast change for various reasons (Harrington, 2012; Kuang

& Cui, 2018). Over time, these cue shifts can have larger impacts on a phonological system, resulting in possible sound changes. Kuang and Cui (2018) describe three possibilities for the time course of a cue shift: (1) shifts in production and perception simultaneously, (2) listeners shift in perception first, then mirror this in production, and (3) change occurs in production first and listeners subsequently are attuned to it in perception. They investigate this by looking at an ongoing change in tense vs. lax register in Southern Yi, in which vowel quality is overtaking phonation as the primary cue. After comparing how this shift is manifested in production and perception patterns in their participants, they found that the shift to formant values occurs first in perception, with production lagging behind, supporting possibility (2) for the time course of cue shifts. Furthermore, they found this change-in-progress to be more advanced in non-high vowels, demonstrating that cue shifting does not need to occur uniformly across and sound system, but can start in a subset of sounds before being initiated in others. This study highlights how investigating acoustic cues as they are produced and perceived and the relationship between these two language modalities can provide rich information not just about how a contrast works in a language but how a language may be changing and the state of possible cue shifting.

Previous research on perception of vowel quantity has begun to show how cues are integrated and weighted for this contrast. In one such study, Pind (1996) looked at Icelandic vowel quantity and found that listeners integrated both vowel duration and quality when identifying long and short vowels. As both vowel duration and quality are correlated with quantity in production, this demonstrates an alignment in the two speech modalities. Behne et al. (1999) found that f0 was an important secondary cue for

Japanese listeners; as f0 differences are prominent in production, this demonstrates another alignment between production and perception. However, the cues that listeners use for a specific contrast are not uniform, they are unique to each language system. This highlights the importance of cross-linguistic studies to understand how a language system uniquely operates; we cannot take perception findings from one language and assume they apply to another simply because they have the same kind of phonological contrast.

There is evidence that the use of acoustic cues in the perception of vowel quantity can be phoneme-specific. Specifically, there is a small body of research looking at the role of vowel quality in the production and perception of long and short high front vowels. Podlipský et al. (2009) investigated the production and perception of /iː/ and /ɪ/ in Czech, a language with phonological vowel length. They found in production, /iː/ was only 30% longer than /ɪ/ while /ɛ, a, oː, uː/ were on average 60% longer than /ɛ, a, o, u/. In perception, they found that listeners relied on vowel quality as much as they did vowel duration for the /iː/-/ɪ/ distinction, while this was not true for the other vowels. Similar findings–that long and short high front vowels are produced with less of a durational difference and listeners use quality more in their perception–is also mirrored in Hungarian (Mády & Reichel, 2007). Studies such as these demonstrate that the way in which cues are correlated with quantity in production and used in perception can, and does at times, vary by vowel within a language.

In the specific case of Norwegian, previous research on how vowel quantity is perceived have been sparse and our understanding of how Norwegian listeners identify long and short vowels is still incomplete. However, there are a handful of studies beginning to form a foundation on which to build ongoing research. It has been shown

that duration is the main cue used by listeners when identifying long and short Norwegian vowels (Nylund & Behne, 1996; van Dommelen, 1999; Behne & Nylund, 2003). However, there is indication that vowel quality is used, though the extent to which is not understood, and there are some indications of vowel-specific patterns (Behne & Nylund, 2003). Furthermore, the duration of the postvocalic consonant has been proven to cause a shift in the perceptual long-short boundary, at least in the case of disyllabic words, as shown in by van Dommelen (1999). Therefore, the further investigation of what cues are used by Norwegian listeners and how they are weighted is warranted to further understand this contrast system.

From here, we can begin to address larger theoretical questions regarding the relationship between speech production and perception via acoustic cues as they exist in Norwegian vowel quantity. For example, can we take the cues that are used and enhanced in production (i.e., Experiment 1 and 2) and predict the cues that are used in perception? What does this alignment tell us about how listeners prioritize cues in perception? For example, if perception is indeed driven by production, we might predict listeners to weight most heavily the cues that are most correlated with category membership in their productions. However, if we encounter a misalignment, what implications does this have for our understanding between production and perception as well as what evidence does this provide to support an ongoing cue re-shifting in Norwegian?

2. RESEARCH QUESTIONS

The main goal of Experiments 1 and 2 was to explore the role of acoustic cues to Norwegian vowel quantity in various types of production. Experiment 1 found that speakers produce vowel quantity with multiple acoustic cues in production including vowel duration, vowel quality, postvocalic consonant duration, and diphthongization of long mid-vowels. In Experiment 2, it was shown that temporal cues are enhanced across all vowels in clear speech to increase intelligibility and salience of the quantity contrast. Acoustic distance between long and short vowels was enhanced for this purpose in four of the six vowels, and long mid vowel diphthongization was not enhanced at all. As it has been proposed in previous literature that when a cue is reliably and systematically present in production, it is often used in perception (Diehl et al., 1994), the next step in the investigation of Norwegian vowel quantity is to see how this principle relates. Furthermore, it is of interest to see whether there is parity in production and perception and the implications this has for the status of acoustic cues in Norwegian as well are our understanding of the production-perception link.

This experiment tests how listeners perceive vowel quantity in Norwegian, and their use of three main cues: (1) vowel duration, (2) vowel quality, and (3) postvocalic consonant closure duration. Specifically: (1) do listeners use all three cues in perceiving vowel quantity, and (2) how are they weighted in relationship to one another. Because all three cues were shown to be correlated with long and short vowels in production, we predict that the three cues will be used in some capacity in perception. We can test this via examining if these variables in the stimuli are significant predictors of participant

quantity categorization. While Experiment 1 confirmed that all three cues are present in the production of vowel quantity, there are patterns in both Experiment 1 and 2 that guide the predicted outcomes of the current experiment. In both Experiments 1 and 2. It was clear that the production and enhancement of vowel quality had vowel-specific patterns and therefore we predict that there might be vowel-specific patterns in perception as well. For example, for vowels that do no exhibit as large of an acoustic difference in perception, listeners might not utilize quality as much in perception. Therefore, responses will be analyzed by-vowel rather than cumulatively.

3. METHODS

3.1 Speaker recordings

The wordlist for the stimuli consisted of six non-word /pVd/ quantitative minimal pairs (see Column A of Table 4.1). For the purpose of the rhyming task described later on, six pairs of real /CVd/ words were recorded, provided in Column B of Table 4.1. The word pairs were produced by a male native speaker of Norwegian in the carrier phrase “jeg sa ___ i går” (I said ___ yesterday). The recordings were made using a Shure WH20 XLR head-mounted microphone in a sound-attenuated booth in the Phonetics Lab at the University of California-Davis and digitized at a 44,100 Hz sampling frequency.

Table 4.1: Orthography for word pairs elicited from the speaker.

Vowel	(A) Stim. pair (non)	(B) Rhyme pair (real)
/i/	pid-pidd	tid (time) - tidd (become quiet)
/u/	pud-pudd	bod (shed) – bodd (lived)
/ɑ/	pad-padd	bad (bath) – ladd (loaded, masc.)
/e/	ped-pedd	fred (peace) - ledd (link)
/ø/	pød-pødd	nød (emergency) – klødd (itched)
/o/	påd-pådd	råd (advice) - rådd (rot)

3.2 Stimuli

To generate the continua for the perception experiment, the vowels were first spliced from the consonant context. Next, they were modified to have a smoothed falling f0 contour beginning at 125 Hz and ending at 100 Hz, decreasing linearly from the start to the end of the vowel. Long vowels in Norwegian can carry pitch accent in Norwegian (Kristoffersen, 2000) and, given that pitch and pitch contour are not of interest in the current study, normalizing the pitch contour eliminates the risk of this biasing listener responses and becoming a confounding factor.

For each vowel pair, a five-step duration continuum was generated in Praat (Boersma & Weenink, 2022) from 70 ms to 150 ms in 20 ms increments; the range of vowel duration is based upon the average durations for long and short vowels as

produced by the speaker and in line with previous research on Norwegian vowel quantity perception (e.g., Behne & Nylund, 2003). Next, a five-step quality continuum was created for each step of the duration continuum using the Praat VocalToolkit (Corretge, 2012) script for blending two sounds. With the first and last steps of the quality continuum being 100% the long or short vowel, the intermediate three steps were varying ratios of the original long and short vowel. The steps are ratios were as follows: (1) 100% short, (2) 75% short and 25% long, (3) 50% short and 50% long, (4) 25% short and 75% long, and (5) 100% long. The two continua created a 5x5 stimulus matrix, similar to designs in previous literature (e.g., Grenon et al., 2019), with 150 total unique vowel stimuli (six vowel pairs, five quality steps, and five duration steps). This grid stimulus design can be seen in Figure 4.1.

In addition to manipulations on the vowel, two postvocalic consonant durations were used: 100 ms and 140 ms. These durations were based on the average postvocalic consonant durations produced by the speaker and confirmed by previous descriptions of vowel quantity production.

Lastly, the isolated vowels were spliced back into their frames, with two versions: one with the “long” postvocalic consonant (140 ms) and one with the “short” (100 ms). All stimuli were normalized for intensity (60 dB).

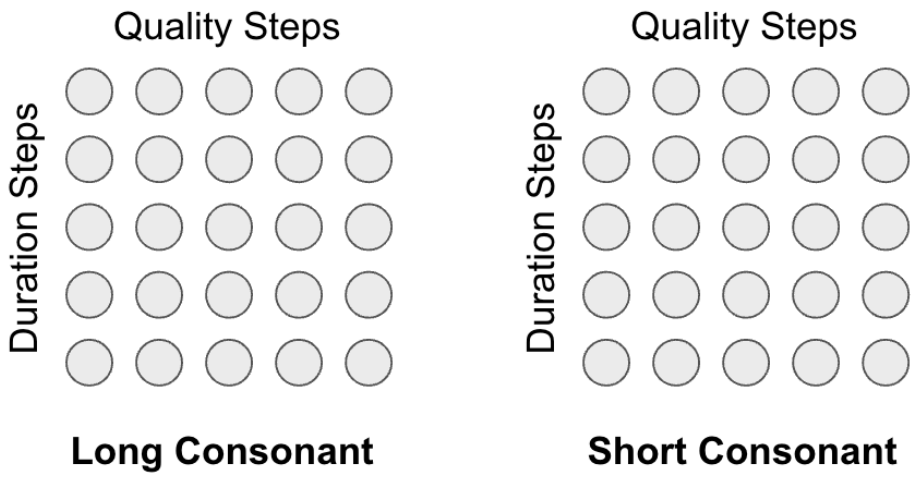

Figure 4.1: Stimuli matrix design for each vowel pair.

3.3 Participants and procedure

53 participants (32 female, 19 male, 2 non-binary; average age = 29.3 years) participated in this experiment at the University of Oslo via online recruitment posts and class visits. All reported being native speakers of Norwegian and all reported speaking at least on other language besides Norwegian. None of the participants reported having any hearing or speech impairments. The study was approved by the UC Davis Institutional Review Board (IRB protocol #1653463-1) and subjects completed informed consent before participating.

Participants were presented with a trial as seen in Figure 4.2. In the first audio file, was the experimental stimulus and word A and B were the rhyme pair as seen in Table 4.1. Participants were asked to listen to the experimental stimulus and determine which of the words (A or B) it rhymed with; this method of categorization was also used by Behne

and Nylund (2003). Whether A or B contained the rhyme word with the long vowel was randomized across the experiment. Each participant completed a random set of 150 trials.

Before the experimental block, participants completed a short block of practice trials containing words not used in the experimental block in order to familiarize them with the task and establish a baseline measure of their ability to complete the experiment. Practice trials contained only unmodified utterances. In the event that a participant was not able to correctly identify the matching pair in the practice trial, their data was not used in the analysis; data from one participant was excluded for this reason.

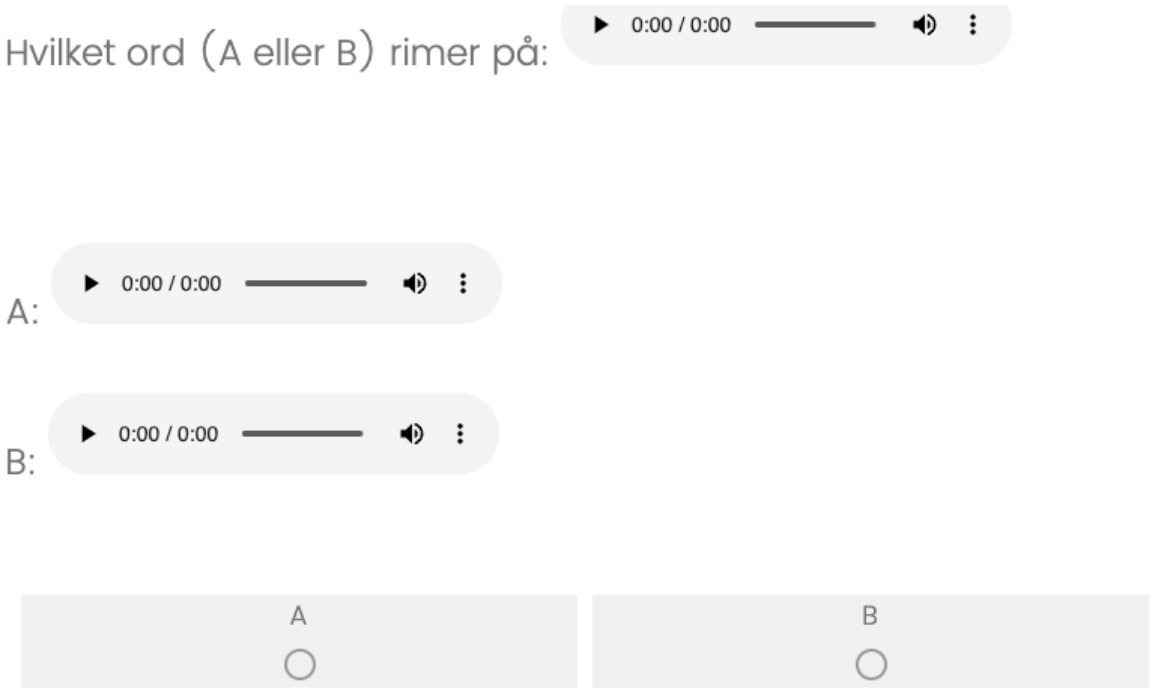

Figure 4.2: Participant view of trials for perception study.

3.4 Statistical analysis

Participant responses were coded for whether stimuli were rhymed with words with a short (=0) or long (=1) vowels and stimuli were coded for Quality (1-5) and Duration (1-5) steps as well as postvocalic consonant length (long or short). A mixed effects logistic regression model was run using the *glmer()* function in the *lme4* package in R (Bates et al., 2015). Fixed effects of the model included Quality (steps 1-5), Duration (steps 1-5), and Consonant (Long vs. Short). Random effects included by-Listener random intercepts and by-Listener random slopes for the main effects (see Equation 1).

$$Response \sim Quality + Duration + Consonant + (1 + Quality + Duration + Consonant \mid Subject) \quad (1)$$

All fixed effects were sum coded. After each model was run, a Wald test was conducted that tests if differences in the estimated coefficients for cues are statistically significantly different (i.e., if listeners weight cues differently). For this test, the rule of thumb is that a value (x) larger than 2 or smaller than -2 indicates the two coefficients are significantly different at a 95% confidence interval (Wheeler, 2016; Ren, 2018).

Separate mixed effect logistic regression models were run for each vowel phoneme. The decision to run vowel-specific models was in order to look at whether fixed effects (Duration, Quality, Consonant duration) were different from zero (e.g., if listeners used these cues at all) rather than the overall average. From these models we can also see the relative cue weights based on the values of the estimated coefficients.

4. RESULTS

4.1 Overall results

Figure 4.3 shows the mean proportion of categorizing the experimental stimulus as long by each Quality step and Consonant length. Figure 4.4 shows the mean proportion of categorizing the experimental stimulus as long by each Duration step and Consonant length. Table 4.2 provides the statistical output from the regression model.

The model showed a significant main effect for all three variables, indicating that all three are used in the overall perception of vowel quantity. The positive coefficient for Duration means that as the vowel's duration increases, listeners are more likely to categorize the vowel as long. For Quality, the positive correlation means that as the quality steps increase (i.e., 100% short to 100% long), listeners are more likely to hear the vowel token as long. The estimated coefficients used by the model can be used to determine relative cue weight (Morrison, 2005; 2007); because both quality and duration had equal steps, we can directly compare their coefficient values. Here, we can see that the model has a larger estimated coefficient for Duration (β=1.030) than for Quality (β=0.814), indicating that overall, listeners weight vowel duration more heavily than vowel quality. This difference in coefficients was found to be significant (x=4.73). Lastly, for Consonant, the positive estimated coefficient means that when the consonant duration is shorter, listeners are more likely to categorize the vowel token as long. As clearly demonstrated in the figures and relatively large coefficient (β=1.249), differences in

postvocalic consonant duration leads to a rather large change in the proportion of long responses, indicating that this is a salient cue for listeners.

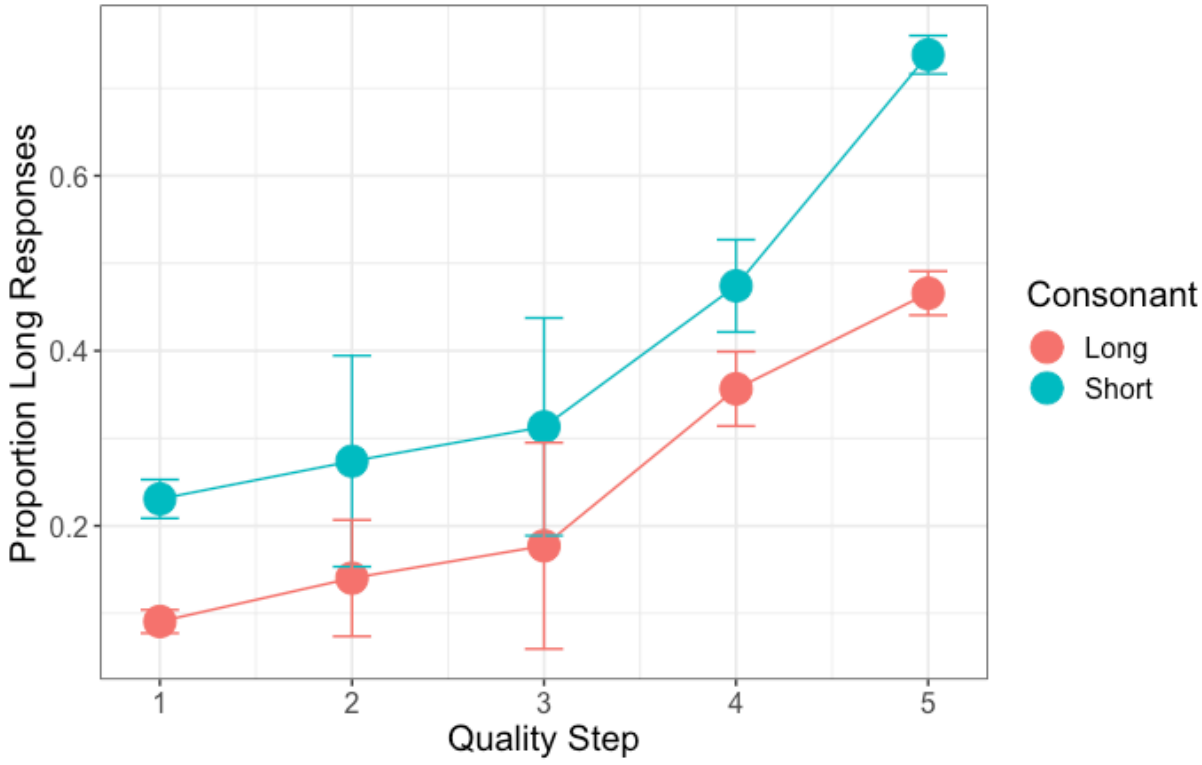

Figure 4.3: Mean proportion of participant identification as long for each quality step ((1) 100% short, (2) 75% short and 25% long, (3) 50% short and 50% long, (4) 25% short and 75% long, and (5) 100% long) by postvocalic consonant type (Long or Short).

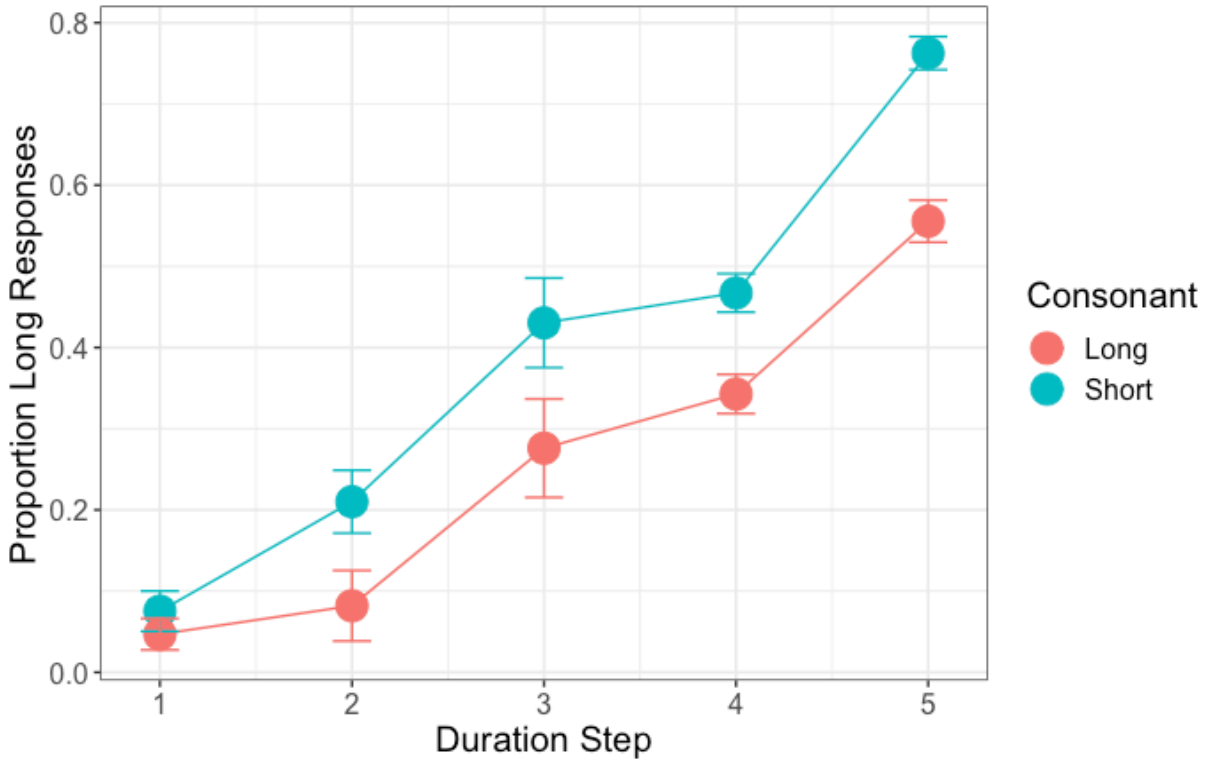

Figure 4.4: Mean proportion of participant identification as long for each duration step (1-5) by postvocalic consonant type (Long or Short).

Table 4.2: Model output for all vowels.

	Est.	Std. Error	*z*	*p*
Intercept	**-6.739**	**0.293**	**-22.930**	**<0.001*** **
Duration	**1.030**	**0.038**	**26.970**	**<0.001*** **
Quality	**0.814**	**0.047**	**17.110**	**<0.001*** **
Cons (Short)	**1.249**	**0.283**	**8.112**	**<0.001*** **

4.2 Vowel-specific patterns

Figures 4.5-10 show the mean proportion of categorizing the experimental stimulus as long by each Quality and Duration step and Consonant length for each vowel. Figure 4.11 shows the coefficient for each cue for each vowel. Tables 4.3-8 provides the statistical output for each model run on individual vowel phonemes, which will be discussed separately below.

The /i/ model showed a significant main effect for all three main effects. This indicates that listeners use vowel duration, vowel quality, and postvocalic consonant duration when distinguishing between long and short /i/. Because the model showed a larger coefficient for Quality (β=0.818) than for Duration (β=0.704), this suggests that listeners weight Quality more heavily for /i/ than Duration. This difference was found to be signiciant (x=4.31). The positive coefficient for Consonant (β=0.682) indicates that listeners are more likely to identify a vowel as long when the consonant is shorter.

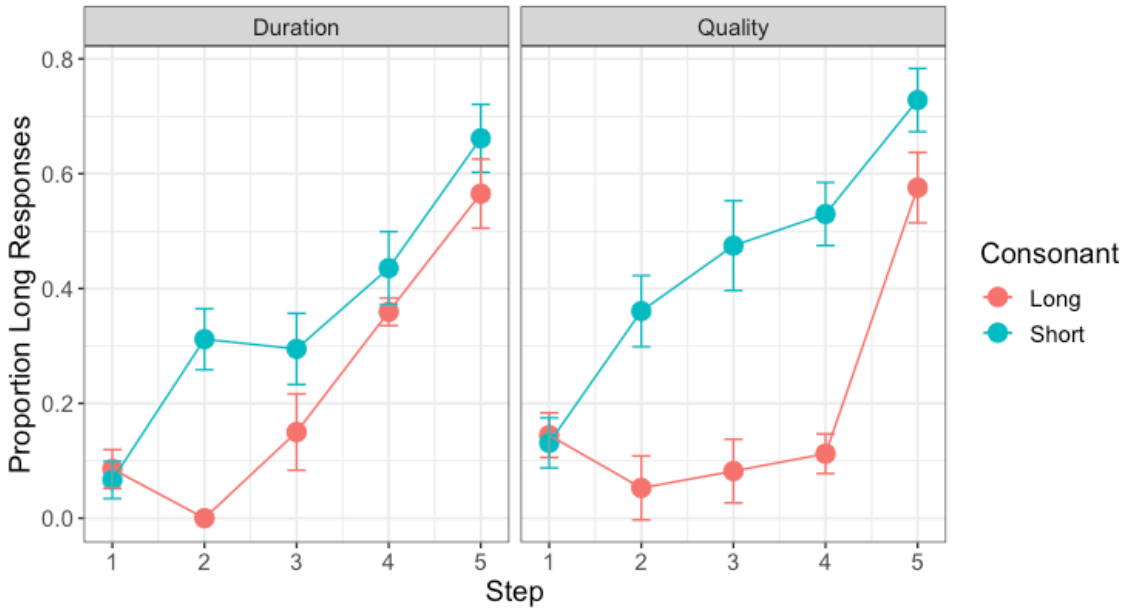

Figure 4.5: Mean proportion of participant identification as long for each duration step (left panel) and quality step (right panel) by postvocalic consonant type (Long or Short) for /i/.

Table 4.3: Model output for /i/.

	Est.	Std. Error	*z*	*p*
Intercept	**-6.439**	**0.516**	**-12.459**	**<0.001*** **
Duration	**0.704**	**0.082**	**8.558**	**<0.001*** **
Quality	**0.818**	**0.086**	**9.512**	**<0.001*** **
Cons (Short)	**0.682**	**0.205**	**3.318**	**<0.001*** **

The /u/ model showed a significant main effect for both Quality and Duration but not for Consonant. This suggests that listeners do utilize both vowel quality and duration when identifying vowel quantity. However, we cannot determine from the model that the effect of consonant is not zero–in the model we can see that the coefficient is one standard

error above zero, suggesting a weak effect that is in line with the other vowels. Furthermore, the larger estimated coefficient for Duration (β=0.916) than Quality (β=0.788) tell us that listers weight vowel duration more heavily than vowel quality for /u/ and this difference was found to be significant (x=4.04).

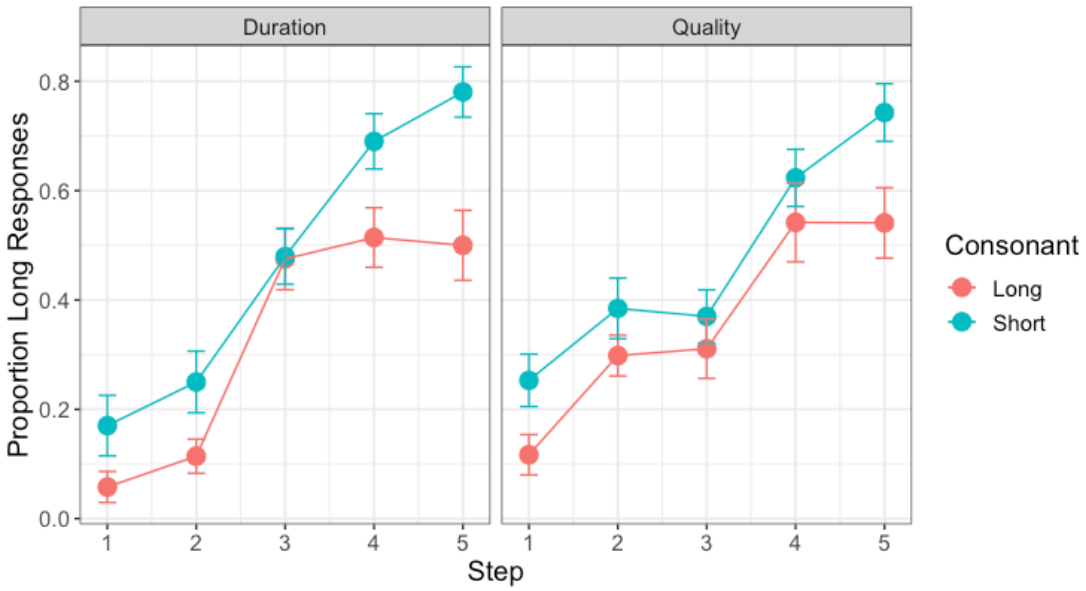

Figure 4.6: Mean proportion of participant identification as long for each duration step (left panel) and quality step (right panel) by postvocalic consonant type (Long or Short) for /u/.

Table 4.4: Model output for /u/.

	Est.	**Std. Error**	***z***	***p***
Intercept	**-6.013**	**0.521**	**-11.524**	**<0.001*** **
Duration	**0.916**	**0.092**	**9.882**	**<0.001*** **
Quality	**0.788**	**0.087**	**9.028**	**<0.001*** **
Cons (Short)	0.267	0.211	1.262	0.207

The /ɑ/ model showed a significant main effect for Duration and Consonant but not for Quality. This means that we cannot say that the effect of quality in the perception of long and short /ɑ/ is not zero and we cannot say definitively whether listeners use this as a cue. The positive coefficient for Consonant (β=0.650) indicates that listeners are more likely to identify a vowel as long when the consonant is shorter.

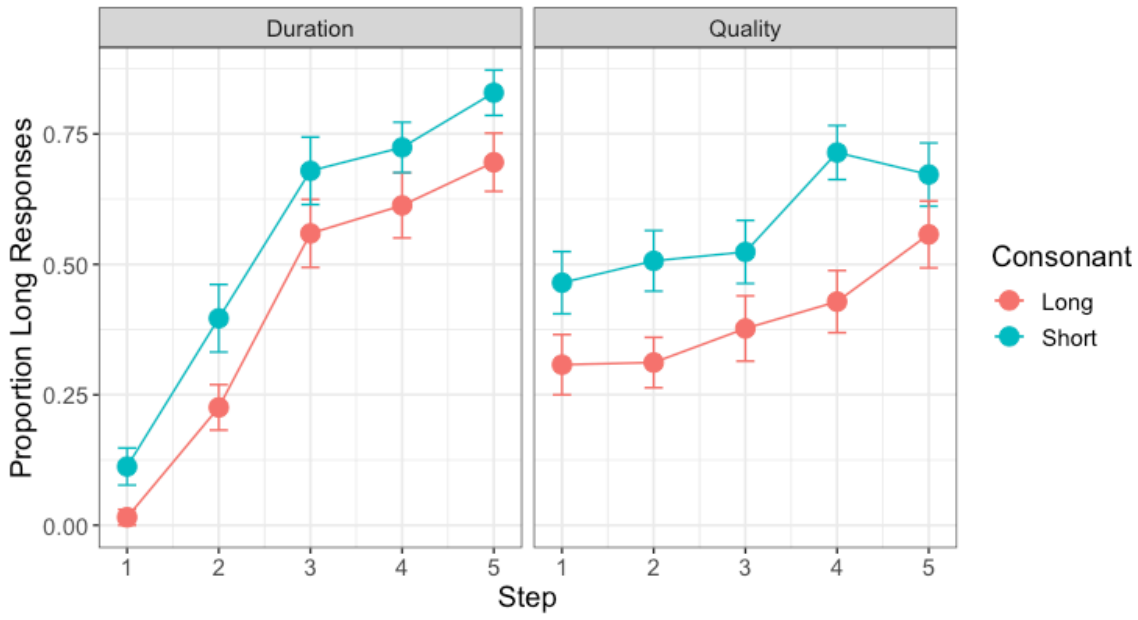

Figure 4.7: Mean proportion of participant identification as long for each duration step (left panel) and quality step (right panel) by postvocalic consonant type (Long or Short) for /ɑ/.

Table 4.5: Model output for /ɑ/.

	Est.	**Std. Error**	***z***	***p***
Intercept	**-6.051**	**0.573**	**-10.543**	**<0.001*** **
Duration	**1.323**	**0.110**	**11.961**	**<0.001*** **
Quality	0.278	0.086	1.342	0.213
Cons (Short)	**0.650**	**0.222**	**2.922**	**0.003** **

The /e/ model showed a significant main effect for all three variables, indicating that listeners use vowel quality, vowel duration, and postvocalic consonant duration in perceiving long and short /e/. Similar to the /i/ model, there was a larger estimated coefficient for Quality (β=2.204) than Duration (β=0.902) indicates that listeners weight a vowel's quality more heavily than a vowel's duration for this vowel. The difference in coefficients was found to be significant (19.02). The positive coefficient for Consonant (β=1.050) tells us that listeners are more likely to identify a vowel as long when the following consonant is shorter.

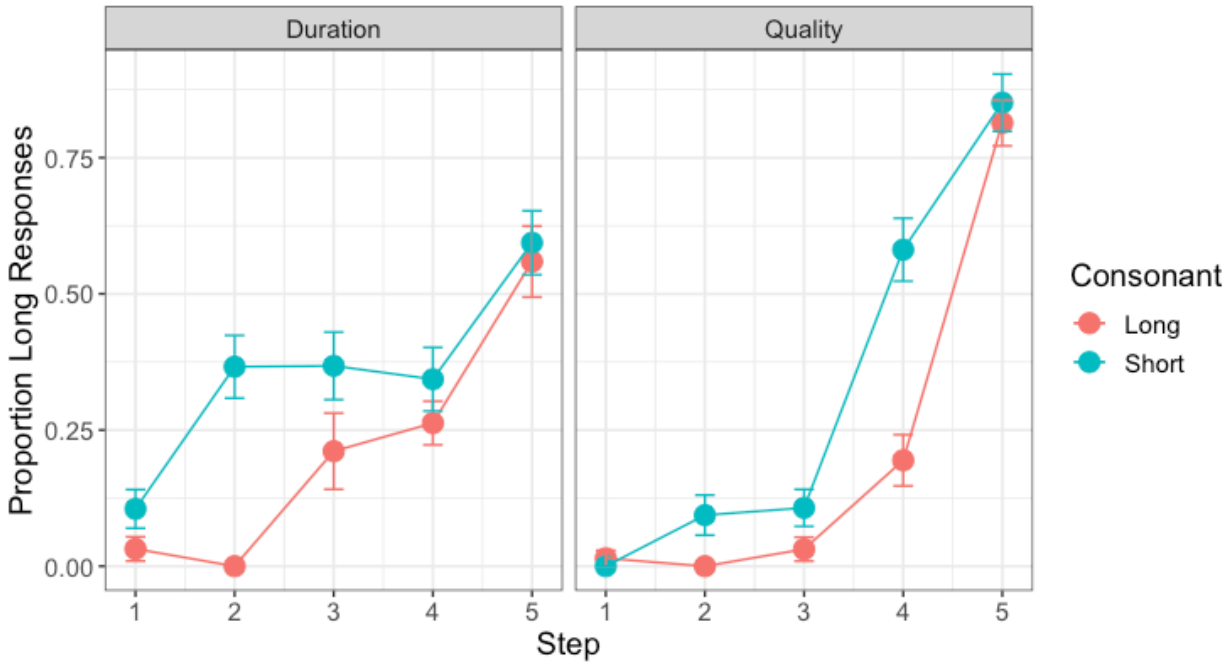

Figure 4.8: Mean proportion of participant identification as long for each duration step (left panel) and quality step (right panel) by postvocalic consonant type (Long or Short) for /e/.

Table 4.6: Model output for /e/.

	Est.	**Std. Error**	***z***	***p***
Intercept	**-12.627**	**1.177**	**10.725**	**<0.001*** **
Duration	**0.902**	**0.125**	**7.171**	**<0.001*** **
Quality	**2.204**	**0.212**	**10.383**	**<0.001*** **
Cons (Short)	**1.050**	**0.294**	**3.561**	**0.003** **

The /ø/ model revealed a significant main effect for all three variables, indicating that listeners do use vowel quality, vowel duration, and the duration of the following consonant in identifying long and short /o/. Furthermore, the larger coefficient for Duration (β=1.661)

than Quality (β=1.095) tells us that listeners weight vowel duration more heavily than vowel quality for /ø/. The difference in coefficients was found to be significant (x=7.63). The positive coefficient for Consonant (β=1.932) also indicates that listeners are more likely to identify a vowel as long when the postvocalic consonant is shorter.

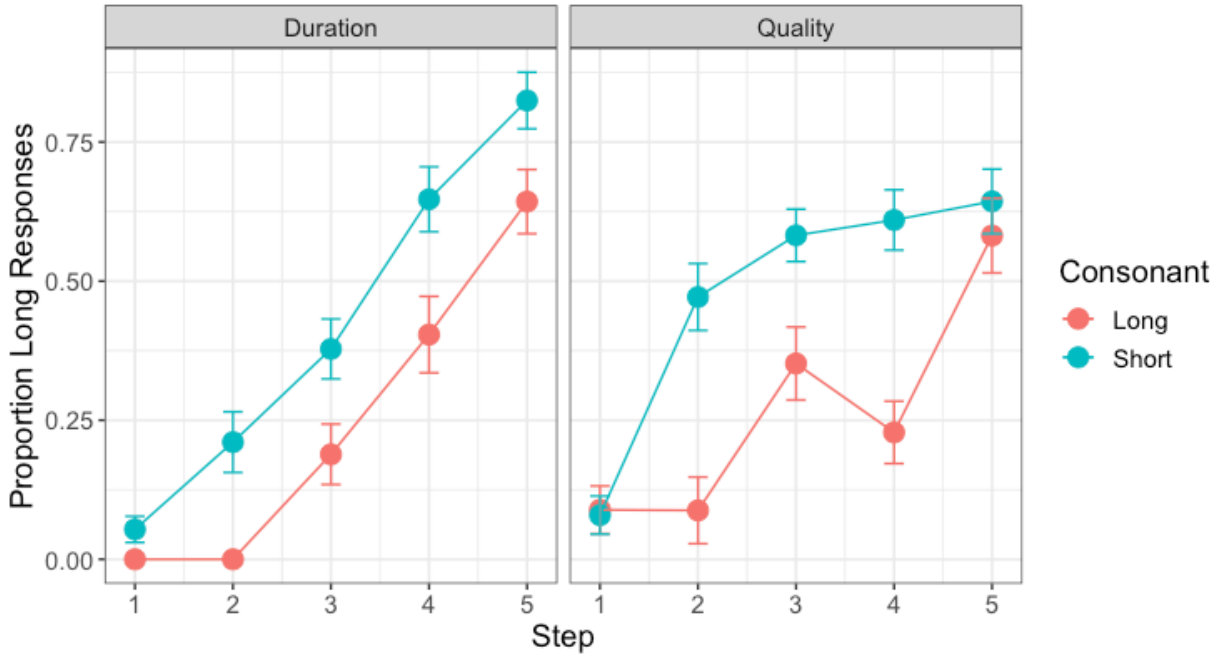

Figure 4.9: Mean proportion of participant identification as long for each duration step (left panel) and quality step (right panel) by postvocalic consonant type (Long or Short) for /ø/.

Table 4.7: Model output for /ø/.

	Est.	Std. Error	*z*	*p*
Intercept	**-10.892**	**0.972**	**-11.199**	**<0.001*** **
Duration	**1.661**	**0.158**	**10.462**	**<0.001*** **
Quality	**1.095**	**0.128**	**8.534**	**<0.001*** **
Cons (Short)	**1.932**	**0.321**	**6.027**	**<0.001*** **

Lastly, the /o/ model showed a significant main effect for all three variables, suggesting that listeners use vowel quality, vowel duration, and postvocalic consonant duration in the perception of long and short /o/. The larger coefficient for Duration (β=1.339) than Quality (β=0.829) indicates that listeners weight vowel duration more heavily than vowel quality and this difference was found to be significant (x=6.74). The positive coefficient for Consonant (β=1.237) tells us that listeners are more likely to identify a vowel as long when the postvocalic consonant is shorter.

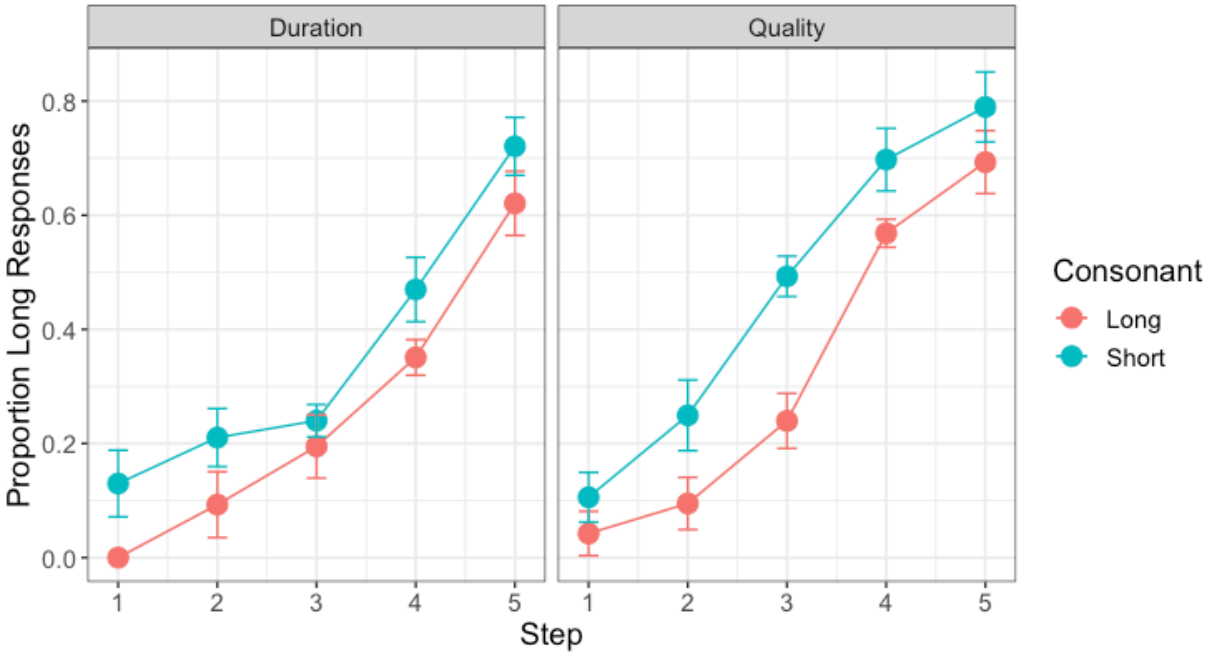

Figure 4.10: Mean proportion of participant identification as long for each duration step (left panel) and quality step (right panel) by postvocalic consonant type (Long or Short) for /o/.

Table 4.8: Model output for /o/.

	Est.	Std. Error	*z*	*p*
Intercept	**-7.170**	**0.756**	**-9.889**	**<0.001*** **
Duration	**1.339**	**0.129**	**10.312**	**<0.001*** **
Quality	**0.829**	**0.106**	**7.755**	**<0.001*** **
Cons (Short)	**1.237**	**0.250**	**4.945**	**<0.001*** **

Figure 4.11 shows the estimated coefficients for each vowel from each of the models; a higher coefficient value indicates a heavier weighting of that cue for that vowel.

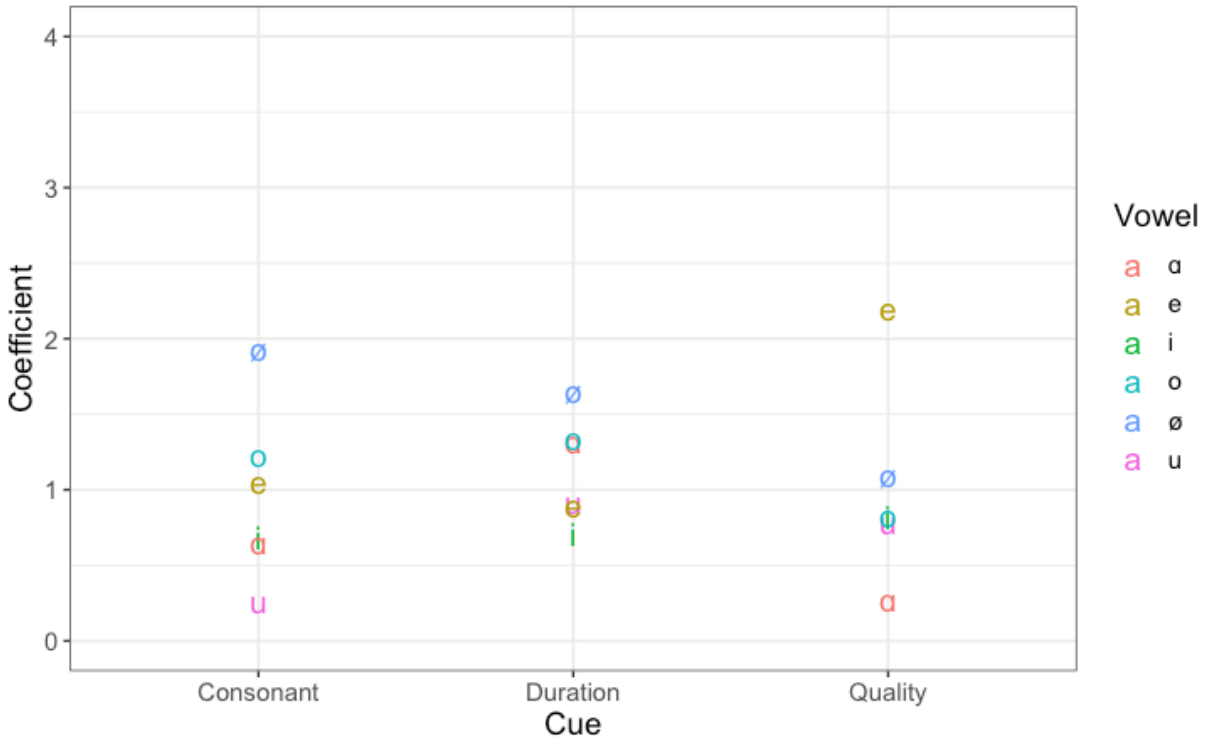

Figure 4.11: The estimated coefficients for each cue and each vowel from models.

5. INTERIM DISCUSSION

5.1 Temporal cues

The first goal of Experiment 3 was to explore how listeners use acoustic cues in the perception of Norwegian vowel quantity. Three acoustic cues were examined: (1) vowel duration, (2) vowel quality, and (3) postvocalic consonant duration.

Through regression analysis of listener responses, the data indicated that listeners use vowel duration to identify the quantity distinction in perception for all six vowels. Furthermore, the positive estimated coefficient calculated by the model indicates that as the duration of the vowel increases, so does the likelihood of listeners identifying the

vowel as long. This is consistent with the presence of vowel duration differences in regular speech and their enhancement in clear speech; as it has been shown that this is an important cue in production, we expected it would play an important role in perception. Because vowel duration is the primary, obligatory cue in vowel quantity, this is not a surprising finding.

The second temporal cue that is investigated is postvocalic consonant duration. Overall, listeners use this cue in perception for five of the six vowels examined. Theories of enhancement discuss how redundant, secondary cues are often used to enhance a primary cue for a contrast and the Duration Ratio Hypothesis specifically describes the mutually enhancing nature of segment duration for VC codas. Given that differences in postvocalic consonant duration were both seen in production of regular speech and enhanced in clear speech, we predicted listeners would use this cue in perception as well. The use of postvocalic consonant duration in both places points to an alignment in production and perception, drawing a link between the two speech modes.

The /u/ model did not reveal a significant effect for postvocalic consonant duration. On the one hand, we can interpret this as listeners not using this cue when perceiving long and short /u/ as much as for other vowels and speculate possible reasons for this. For example, it might be that the other two cues of vowel quality and duration are more prominent for listeners when compared to postvocalic consonant duration, leading them to not use it in perception. On the other hand, Figure 4.6 clearly illustrates a higher likelihood of participants categorizing a vowel token as long when it is followed by a shorter consonant. Thus, it is reasonable to assume that with simply more data points, this effect would indeed be significant in the model. While it is interesting to speculate

why listeners would not use postvocalic consonant duration in perceiving long and short /u/, we will not dismiss the likelihood that this is more or less a random finding.

5.2 Spectral cues

Based upon the stance that when cues are reliably and systematically correlated with a contrast in production, they will be used in perception (Diehl et al., 2004), we predicted that vowel quality would be used in perception by listeners. This is motivated from the fact that in Experiment 1, it was shown that all six vowels had significant differences in vowel quality between long and short vowels, even though the way in which they differed varied. Therefore, we predicted that vowel quality would be used in perception for all six vowels. This was the case for five of the six vowels, but not for /ɑ/. Here, we could not verify that listeners use vowel quality. Though there are spectral differences between long and short /ɑ/ in production, this difference was not enhanced (although still present) in clear speech. The lack of enhancement raises the possibility that spectral differences are a consequence of articulation (i.e., longer vowel durations allow more peripheral articulations) rather than part of the underlying representation. This possibility is further supported by listeners' failure to reliably utilize vowel quality in the perception of long and short /ɑ/. The misalignment between production and perception raises interesting questions about the nature of quantity for this vowel and the vowel system in general. Could this be evidence for an incomplete cue shifting in Norwegian? For example, while the integration of vowel quality differences between long and short vowels is complete in production, it might not be for perception where vowels for which listeners utilize quality

more heavily than duration (i.e., /i/ and /e/) have led the cue shift while low vowel /ɑ/ is lagging behind. Of course, a definite conclusion cannot be made from this experiment alone, but this point would be an interesting point for future research.

In addition to whether listeners used vowel quality in speech perception, cue weighting was also a central point of investigation in this study. The estimated coefficient calculated by each model was used to estimate the relative weight and ordering of vowel quality and duration; because postvocalic consonant duration had only two levels while the other two variables had five, we are not able to directly compare its estimated coefficient in the same way. In the models for /u, o, ø/, vowel duration was weighted more heavily than vowel quality. This is not surprising and falls in line with previous cue weighting studies which show that the primary cue for a contrast is typically weighted the heaviest. However, for /i/ and /e/, these models diverged from this pattern in that vowel quality had a higher estimated coefficient, suggesting that listeners actually weight vowel quality more heavily than vowel duration for these two vowels. In Experiment 1, the two vowels /i/ and /e/ were highlighted in Section 5.2 for having exceptionally large acoustic distances between the long and short vowels. As stated previously, the /i:/-/ɪ/ contrast (both qualitative and quantitative) has been the topic of much previous research, largely due to its tendency cross-linguistically to be an especially salient contrast (e.g., Kim et al., 2020). Consequentially, this pair is often used in studies on acoustic cues and weighting in both production and perception. In terms of these two vowels in perception, specifically of vowel quantity, there is cross-linguistic evidence that this pair often includes differences in vowel quality that are more pronounced than other vowels (for Hungarian: Mády & Reichel, 2007; for Czech: Podlipský et al., 2009). The data from these studies is

presented in the context of cue shifting in these languages, during which the primary cue for quantity in some parts of the vowel system is becoming vowel quality rather than duration. Because the importance of vowel quality is in only part of the vowel system, this supports an argument that this change is in progress.

Returning to the data presented in Experiment 1, these two vowels do have a larger acoustic difference between long and short vowels than other vowels. Therefore, it is possible that this heightened distinctiveness in production carries over to perception: listeners have experience with hearing long and short /i/ and /e/ produced with particularly large quality differences and have, thus, begun to rely more heavily on vowel quality for these two vowels because of that.

5.3 General remarks

The original goal of this experiment was to establish how the following cues are used in the perception of Norwegian vowel quantity: (1) vowel duration, (2) vowel quality, and (3) postvocalic consonant duration. From the data here, we can see that all three cues play are used in perception, although there are vowel-specific patterns. Vowel duration is used by listeners for all six vowels, supporting that it is the primary acoustic cue for this contrast. Vowel quality is used by listeners for five of the six vowels: the one low vowel /ɑ/ was the exception, reminiscent of Experiment 2 where speakers did not enhance vowel quality for this vowel either. When perceiving long and short /i/ and /e/, listeners weight vowel quality more heavily than vowel duration; this mirrors Experiment 1 where these two vowels had particularly large acoustic distances between long and short vowels, which may begin to explain the pattern in perception. The duration of the postvocalic consonant is used by

listeners for five of the six vowels: listeners do not use this cue for long and short /u/, although an exact explanation for why this is the case is not clear.

CHAPTER 5 – GENERAL DISCUSSION

1. RESTATEMENT OF PROBLEM

The goal of this dissertation was to explore the link between phonetic variation and systems of phonological contrast, focusing on spectral and temporal cues in the production and perception of Norwegian vowel quantity. It has been well-documented that phonemic category membership is rarely defined by a single acoustic cue. For example, the seemingly simple contrast of stop consonant voicing in English is signaled by up to sixteen different acoustic cues (Lisker, 1986). Despite a number of theories dealing with the existence of multiple cues in phonological contrast, there are still many underaddressed questions about how secondary acoustic cues function. Thus, a central goal of this dissertation was to explore the complexity of deceptively simple-seeming phonological contrasts via the role of primary and secondary cues. Moreover, this dissertation aimed to add a meaningful contribution to a growing body of literature aimed at deepening our understanding of these concepts.

Another goal of this dissertation was to explore the relationship between speech production and perception, two speech modalities that have traditionally been examined independently. However, understanding their relationship and where they align or misalign helps to create a more comprehensive model of representation and linguistic knowledge (Casserly & Pisoni, 2010). Correlations between production and perception have been found, for example in studies focusing on contextualizing cues. In one such study, the degree of coarticulatory vowel nasalization in speech production correlates with

sensitivity thereto in perception (Zellou, 2017). Yet, this area needs to be investigated further and this dissertation examined these questions through the lens of acoustic cues correlating with the Norwegian vowel length distinction. Do we see that acoustic correlates in production are mirrored in perception? Or, alternatively, if there is a misalignment the in acoustic correlates found in production and the cues used in perception, what does that mean? Could this point to an ongoing cue shift in Norwegian?

Prior works that have provided production-based descriptions of vowel quantity in Norwegian have varied in whether they include the qualitative difference between long and short vowels. One group of descriptions state that only temporal cues (i.e., vowel and postvocalic consonant durations) are correlated with quantity in production (Behne et al., 1996; van Dommelen, 1999) while others describe long vowels as being more peripheral than short vowels (Kristoffersen, 2000). In perception, the role of primary and secondary cues has not been clearly defined in Norwegian. Previous studies examining the role of spectral and temporal information in quantity perception found evidence that quality might be used by listeners, but only for some vowels within the experimental subset (Nylund & Behne, 1996; Behne & Nylund, 2003). Van Dommelen demonstrated that the duration of the postvocalic consonant affected the categorization of long and short vowels via moving the perceptual boundary between quantities. The precise role of secondary cues in the production and perception of Norwegian vowel quantity has not been clearly defined or extensively investigated and there are ambiguities and uncertainties remaining. Thus, this dissertation had several aims.

With respect to production, this dissertation served two main purposes. First, to be a comprehensive acoustic description of the acoustic correlates of vowel quantity as they

are produced by Norwegian speakers and whether or not there is uniformity across the vowel system. Second, this dissertation investigated the way in which these acoustic cues are enhanced in clear speech produced for the clarification of misheard speech; this dissertation was the first study to examine this specific context. The ways in which speech is enhanced will be examined within the context of what this can tell us about the nature of vowel quantity in Norwegian and what speakers deem to be representative of a phonemic category. With respect to perception this dissertation examined how listeners utilize both temporal and spectral cues in perceiving long and short vowels and how they adapt their use of secondary acoustic cues when a primary cue is no longer informative? This dissertation explored listeners' weighting of vowel duration, vowel quality, and postvocalic consonant duration in identifying long and short vowels as well as vowel-specific perceptual strategies.

The findings of the three experiments in this dissertation are discussed with the goal of understanding the role of sub-phonemic information in the phonetic grammar of Norwegian. How the phonetic realization of quantity in production and how its perceived is also discussed in comparison to cross-linguistic patterns established in previous literature. Furthermore, the relationship between production and perception in speech communication is examined via the alignment or misalignment of how the acoustic correlates of quantity in production are used by listeners in perception.

2. SUMMARY OF FINDINGS

1.1 Production

In Experiment 1, the goal was to provide an acoustic analysis of the realization of Norwegian vowel quantity in a set of six long-short vowel pairs for: /i/, /u/, /ɑ/, /o/, /e/, /ø/. Specifically, do Norwegian speakers use multiple cues in producing quantity contrasts? What are these cues and are they uniform across the vowel system? Four acoustic cues were examined: (1) vowel duration, (2) vowel quality, (3) postvocalic consonant duration, and (4) degree and direction of spectral movement in diphthongized long vowels.

Of two temporal cues (1) and (3), both were shown to be correlated with quantity in production. Previous literature described long vowels as being anywhere from 1.4 to 3.3 times as long as short vowels (Fintoft, 1961; Vanvik, 1972; Payne et al., 2017). Participants in Experiment 1 produced long vowels approximately 2.87 times as long as short vowels, in line with the range provided by previous literature. There was not a different by-vowel in durations, even though previous cross-linguistic research has found that low vowels are often intrinsically longer than high vowels, a consequence of their articulation requiring the jaw to open further (Solé & Ohala, 2010). For postvocalic consonant duration, previous literature described consonants after short vowels as being anywhere from 1.0 to 1.8 times the duration of consonants after long vowels. The participants in Experiment 1 produced consonants after short vowels 1.4 times as long as those after long vowels, falling within the previously established range.

Of the two spectral cues (2) and (4), both were also shown to be correlated with quantity in production. It has been well established in the literature on vowel quantity that

differences in quality in long-short vowel pairs are common cross-linguistically (Maddieson, 1984). In the six vowels in the current study, all had a significant acoustic difference between long and short vowels along at least one formant dimension (F1 or F2). Of the six vowels, two vowels had an especially large acoustic distance in the long-short pair: /i/ and /e/. With respect to spectral movement indicating diphthongization of long mid vowels, long /e, o, ø/ were produced with a significantly higher amount of spectral change, specifically with movement toward the center of the vowel space.

Table 5.1: Summary of Experiment 1 findings (✓ = produced differently, × = not produced differently).

Cue	i	u	ɑ	e	ø	o
Vowel Duration	✓	✓	✓	✓	✓	✓
Vowel Quality	✓	✓	✓	✓	✓	✓
Postvocalic Consonant Duration	✓	✓	✓	✓	✓	✓
Diphthongization	×	×	×	✓	✓	✓

1.2 Enhancement

Experiment 2 aimed to investigate the way in which speakers adjust their speech, manipulating acoustic-featural distinctions correlated with the vowel quantity contrast when trying to clarify misheard speech. For example, if the speaker produces a long vowel that the interlocutor, in turn, misinterprets as a short vowel, how will the speaker adjust their speech when correcting them? Previous research provides evidence that speakers

make local, targeted adjustments in clear speech. For example, when an interlocutor misunderstands a stop's voicing (e.g. perceiving /p/ and /b/), speakers hyperarticulate the VOT on that segment in clarifying (Schertz, 2013). This experiment sought to explore if this is the case for Norwegian vowel quantity and what sort of targeted adjustments are made for this contrast.

For vowel duration, the relative durations of long and short vowels were analyzed through a duration ratio, taken by dividing the duration of the long vowel by the duration of the short vowel; thus, a larger ratio indicates a larger relative difference. In correction trials, the duration ratio increased for all vowels, indicating that speakers were enhancing the durational differences between long and short vowels in clear speech. Furthermore, the average duration of short vowels decreased. Because overall increased segment duration is documented as a global feature in clear speech, the shortening of short vowel durations shows speaker-controlled, contrast-focused adjustments. For postvocalic consonant duration, a similar ratio was calculated, with consonant durations after short vowels divided by those after long vowels and a larger ratio was indicative of a larger relative duration difference. This ratio did increase in clear speech compared to regular speech, indicating that postvocalic consonant durations are enhanced by speakers in clear speech as well. Similar to what was observed with vowel duration, the postvocalic closure duration after long vowels also shortened in correction trials compared to initial utterances, indicating this is a contrast-focused adjustment.

In addition to temporal cues, both vowel quality and degree of spectral movement in clear speech were investigated. As a global marker of clear speech is an expanded vowel space (Bradlow, 2002), enhancement of quality differences between long and short

vowels was investigated via Euclidean distance within vowel pairs; a larger Euclidean distance in clear speech indicates that the two vowels are being produced further apart in the vowel space and, thus, more acoustically differently. It was found that speakers produced vowel pairs with a larger Euclidean distance for four of the six vowels: there was not a significant difference in the distance between long and short /ɑ/ across regular speech (i.e., initial utterances) and clear speech. This indicates that while listeners enhance quality differences for the other five vowels, they do not for /ɑ/, demonstrating vowel-specific patterning in contrast enhancement. Additionally, the acoustic distance between long and short /o/ decreased rather than increased, also indicating that listeners did not enhance vowel quality differences for this vowel either. Furthermore, for the three mid vowels, the degree of spectral movement was tested between utterance types to see if speakers enhanced diphthongization of long mid vowels in clear speech. It was found that there was no significant difference in degree of movement between regular speech and clear speech for long mid vowel, indicating that listeners did not enhance this feature in clarifying vowel quantity.

Table 5.2: Summary of Experiment 2 findings (✓ = enhanced, × = not enhanced).

Cue	**i**	**u**	**ɑ**	**e**	**ø**	**o**
Vowel Duration	✓	✓	✓	✓	✓	✓
Vowel Quality	✓	✓	✓	✓	✓	✓
Postvocalic Consonant Duration	✓					
Diphthongization				×	×	×

1.3 Perception

The goal of Experiment 3 was to explore how listeners use acoustic correlates of Norwegian vowel quantity in perception. It was shown in Experiment 1 that all four acoustic cues investigated were correlated with vowel quantity in production. Therefore, Experiment 3 sought to investigate whether these cues from production were used by listeners in perception and how they were weighted. Furthermore, as vowel-specific patterns were seen in both Experiments 1 and 2, finding whether similar vowel-specific patterns in perceptual strategies existed was another goal of Experiment 3. The findings of Experiment 3, along with those from Experiments 1 and 2 also give insight into the link between speech production and perception.

Regression analysis of listener quantity categorizations showed that listeners use vowel duration in perceiving long and short vowels for all six vowels in the experiment: as vowel duration increased, so did the likelihood that a listener would categorize the vowel as long. Given that vowel duration is the primary cue for vowel quantity, it was expected that the cue would play an important role in perception. The second temporal cue that was investigated was postvocalic consonant duration, and listeners reliably used this in quantity perception for five of the six vowels; they did not use this to the same degree when distinguishing between long and short /u/. Given the mutually enhancing nature of vowel and consonant duration in VC sequences (i.e., the Duration Ratio Hypothesis), listener reliance on postvocalic consonant closure duration was predicted.

Vowel quality was reliably used by listeners for five of the six vowels (all but /ɑ/). It was predicted that quality would be predictive of listener responses for all six vowels,

based on the patterns seen in production, where long and short vowels were produced with distinct spectral qualities for all six vowels; this is a misalignment between production and perception. However, listeners not reliably using quality in distinguishing between long and short /ɑ/ patterns with how speakers enhanced vowel quantity: while they produced long and short vowels with more different qualities for the other vowels, they did not for /ɑ/.

In addition to whether listeners used cues, we also looked at the relative cue weights of vowel duration and quality. Of the five vowels where listeners relied on vowel quality when determining a vowel's quantity, vowel duration was more heavily weighted than quality for three vowels /u, o, ø/ while quality was more heavily weighted than duration for two vowels /i, e/. In Experiment 1, it was found that these two vowels were produced with a larger acoustic distance in the long-short pair than other vowels, suggesting that this heightened distinctiveness in quality carried over to perception.

Table 5.3: Summary of Experiment 3 findings (✓ = used, × = not used).

Cue	**i**	**u**	**ɑ**	**e**	**ø**	**o**
Vowel Duration	✓	✓	✓	✓	✓	✓
Vowel Quality	✓	✓	×	✓	✓	✓
Postvocalic Consonant Duration	✓	×	✓	✓	✓	✓

1.4 General remarks

This dissertation investigated the role of four acoustic cues in the production, enhancement, and perception of Norwegian vowel quantity: (1) vowel duration, (2) vowel quality, (3) postvocalic consonant duration, and (4) diphthongization of long mid vowels. In production of regular speech (Experiment 1), it was shown that all four cues under investigation were correlated with vowel quantity in Norwegian, suggesting they all play a role in defining this contrast in this language. Their role was further explored in clear speech (Experiment 2), where the ways in which these cues are enhanced by speakers was investigated. It was shown that vowel and postvocalic consonant duration was enhanced across the board while degree of spectral movement in long mid vowels was not, indicating that not all cues present in production are enhanced in clear speech. Furthermore, vowel-specific patterns emerged when the difference in vowel quality was not enhanced for /ɑ/ thought it was for the other five vowels.

In perception (Experiment 3), how these cue (1)-(3) are used by listeners was investigated. It was found that vowel duration is used across the board by listeners in perceiving long and short vowels. However, vowel-specific patterns emerged again in terms of how listeners used vowel quality and postvocalic consonant duration: listeners do not use vowel quality in perceiving long and short /ɑ/ nor do they use postvocalic consonant duration in perceiving long and short /u/. Furthermore, in perceiving /i/ and /e/, they weight vowel quality more heavily than duration. This illustrates that not only vowel-specific patterning in if listeners utilize cues but also in how much they rely on independent acoustic cues.

3. COMPARING NORWEGIAN AND CROSS-LINGUISTIC PATTERNS

3.1 Production

Research looking into the acoustic correlates of vowel quantity in production has described patterns in a number of languages, displaying common cross-linguistic traits as well as more language-specific patterns. For example, while vowel quality (for Swedish, Behne et al., 1997) and postvocalic consonant duration (for Icelandic, Pind, 1996; for Swedish, Behne et al., 1999) are commonly seen cross-linguistically, cues such as f0 (for Japanese, Kinoshita et al., 2012) are not. Thus, considering how the patterns found in Norwegian fit into established accounts from other languages helps to provide insight into vowel quantity as a phonological contrast as well as the complexity and language-specificity of language contrasts in general.

Cross-linguistically, a number of additional cues to vowel duration have been noted in production. One of the most commonly cited secondary cues is vowel quality in which long and short vowels are produced with differences in spectral characteristics measurable in the first two formants. Maddieson (1984, p. 129-130) provides an account of 331 languages, of which 56 languages had vowel quantity and 17 languages were shown to have differences in vowel quality between long and short vowels. As shown in Experiment 1, Norwegian speakers reliably produce long and short vowels with different qualities too, although the degree and direction of difference varies vowel-to-vowel. In addition to differences in vowel quality via position in the F1/F2 plane, Norwegian diphthongize long mid vowels, producing them with increased spectral movement

compared to long non-mid vowels and short vowels. Specifically, this diphthongization is done in a centralizing direction. There are descriptions of a similar process, diphthongizing long vowels, in other languages. One example is the diphthongization of long /ɑ/ in the variety of the German dialect Kölsch spoken in Dane County, Wisconsin (Siefert, 1963). Furthermore, the diphthongization of only a subset of the vowels demonstrates the nuanced nature of vowel quantity in Norwegian.

Acoustic differences realized on adjacent segments via compensatory lengthening or shortening is another common acoustic cue correlated with vowel quantity. In fact, mutual enhancement of duration in VC segments is extremely common (Kingston & Diehl, 1994). As such, we see a similar pattern in Norwegian to other languages: consonants produced after short vowels are reliably and systematically produced with a longer duration than consonants produced after short vowels. Similar compensatory patterns are seen in Swedish (Elert, 1964), where listeners also regularly lengthen consonants after short vowels compared to long.

In sum, the patterns seen in Norwegian both coincide with common cross-linguistic traits and add something new to the literature on long and short vowels. In addition to the primary cue of vowel duration, the production of long and short vowels with reliably different qualities and variation in the postvocalic consonant duration is in line with what we see in a number of other languages (i.e., Pind, 1996; Behne et al., 1999). Yet, the diphthongization of specifically long mid vowels is something that has not been, to our knowledge, attested as a marker of phonological quantity in another language. This adds to the list of secondary acoustic cues for long and short vowels and demonstrates both the complexity of this phonological contrast but also the language-specific nature of it.

3.2 Perception

A body of work has attempted to describe which of the secondary cues found to correlate with quantity in production are actually used by listeners in perception. Specifically, there has been a goal of understanding how secondary cues are used by listeners in identifying the vowel quantity contrast. Cross-linguistically, the importance of secondary cues both on the vowel and on adjacent segments (i.e., the following consonant) has been demonstrated. These cues include segmental context (Tranmüller & Krull, 2003), dynamic f0 (Lehiste, 1976; van Dommelen 1993; Lippus et al., 2013), and spectral differences between long and short vowels (Abramson & Ren, 1990; Sendelmeier, 1981). In addition to serving as a description of the secondary cues that are used in the perception of vowel quantity, these accounts also provide evidence of the multidimensional nature of not just this contrast, but of phonological contrast in general.

The data from Experiment 3 add to these previous findings in showcasing the complexity of vowel quantity perception and vowel quantity as a multi-dimensional contrast. Similar to other languages, Norwegian listeners were shown to utilize both temporal and spectral cues on the vowel and adjacent segments in categorizing long and short vowels. The use of vowel quality in perception is relatively common in quantitative contrasts (for Icelandic, see Pind, 1996; for German, Thai, and Japanese, see Lehnert-LeHouillier 2010). However, there are some vowel-specific patterns that emerge. First, vowel quality is used by listeners in identifying long and short vowels for all but /ɑ/–such vowel-specific use of cues is attested (for Swedish, see Behne et al., 1997) but not

common cross-linguistically. Furthermore, listeners do not appear to utilize postvocalic consonant duration in the categorization of long and short /u/.

The finding that listeners did not use vowel quality when distinguishing between long and short /ɑ/ is interesting in light of findings from a very closely related Scandinavian language: Swedish. Behne et al. (1997) looked at the perceptual weight of vowel duration and the first two formant frequencies in vowel quantity perception in Swedish. Looking at /i, o, ɑ/, the researchers found that while listeners, of course, used vowel duration in determining vowel quantity for all three vowel pairs, the only one they used vowel quality for was /ɑ/. Norwegian and Swedish are two closely related, mutually intelligible languages that have a number of similar phonetic and phonological qualities; of these, one is that both languages have vowel quantity that includes differences in quality in long and short vowel pairs. Despite how closely related these two languages are, the inverse relationship between the patterns we see in Swedish (only using quality for /ɑ/) and Norwegian (using quality for all vowels except /ɑ/) is extremely interesting.

3.3 Vowel height and salience of vowel quality

Between Experiments 1 and 3, vowel quality has been shown to play a stronger role in the production and perception of long and short /i, e/ than the other vowels in the set. Specifically, speakers produced long and short /i/ and /e/ further apart in the vowel space than other vowels, and these were the only two produced with significant differences along both F1 and F2. In perception, listeners weighted vowel quality more heavily than vowel duration in the perception of these as well.

This pattern of listeners using and weighting cues differently in the perception of high front vowels–specifically in weighting vowel quality more heavily–is attested cross-linguistically. For example, similarities can be drawn between the data presented here and other languages such as Czech, where quantity classification of /iː/-/ɪ/ entails listeners relying more on vowel quality than duration (Podlipský et al., 2009); for the other vowel pairs in Czech, duration remains the most heavily weighted cue. The special status of front high vowels in quantity perception is shown in other languages as well (for example, Hungarian: Mády & Reichel, 2007). In both Czech and Hungarian, the researchers offer an acoustic explanation for this pattern, shown in production patterns in their respective languages. Specifically, in Mády & Reichel (2007) the researchers claim that the duration-based quantity contrast in Hungarian has "eroded" somewhat, given that speakers produce long and short /i/ in Hungarian with less distinct differences in duration. Thus, the researchers explain, duration is less reliable as a cue for this phoneme and listeners upweight vowel quality to compensate for this. But how does this square with the findings here on the production of Norwegian vowel quantity? To put it simply, this sort of explanation cannot generalize to Norwegian. There is no current evidence that either /i/ or /e/ is being produced with a smaller relative duration difference; as seen in Experiment 1, long and short vowels for all six pairs were produced with roughly the same long-to-short ratio.

Though the patterns look, on the surface, to be similar, there must be another explanation for listeners weight vowel quality more heavily for some vowels than others. Given that a lack of reliability of the primary cue, duration, is not to blame, we might turn our attention to the possibility that this pattern could come from vowel quality being extra

salient and reliable. As previously stated, long and short /i/ and /e/ are produced by speaker with a larger relative acoustic distance between them compared to the other vowels in the experiment. Furthermore, these two vowel pairs were the only ones to have differences in both F1 and F2 values. Thus, it is possible that this larger relative qualitative difference between long and short vowels for these might be more salient for listeners and through their experience with hearing them produced this way, listeners have learned that vowel quality is a more reliable cue here. This could in turn lead to the pattern seen in the data, where listeners weight quality more heavily than duration. Now, the origin of this larger difference in production is unclear and would require further research to uncover.

From the current data, it is not possible to make a definite claim about why listeners have a different cue ordering for high vowels, further research is required to reach this point. However, the data can offer some possible starting points for this research and allow speculation on our part and consideration of how the patterns in Norwegian compare to others seen cross linguistically.

4. CLEAR SPEECH

Lindblom's H&H (hypo- and hyperarticulation) Theory introduces an account of variation that is based upon communicative context, illustrating speech as a dynamic and adaptive act. In communicative contexts where conditions favor the listener, speakers are said to produce hypospeech marked by quicker speaking rates and more phonetic reduction. Contrastively, when communicative contexts do not favor the listener, speakers are said

to produce hyperarticulation, marked by a number of phonetic adjustments aimed at maximizing intelligibility.

Hyperarticulation has become an umbrella term for various adaptations in speaking. A body of literature has observed clear speech as a mode of speaking, uncovering a handful of global characteristics. Suprasegmental characteristics include speech that is 5 to 8 dB louder than conversational speech (Picheny et al., 1986), slower speaking rates as measured in words per minute (Picheny et al., 1986), and a larger range of f0 (Bond et al., 1989; Summers et al., 1988). More segment-focused effects have also been found in clear speech; for example, vowels in cellar speech are often produced with an expanded vowels space and increased duration (Chen, 1980; Picheny et al., 1986; Moon & Lindblom, 1994; Bradlow, 2002; Krause & Braida, 2004).

One important question within research on hyperarticulation and clear speech is how speakers dynamically adjust their productions to suit specific communicative challenges. Specifically, do speakers make local adjustments to their articulations that are specifically targeted at sources of phonological confusion? In the particular communicative context of misheard speech, targeted adaptation accounts claim that speakers make segmentally-targeted adaptation specifically to the phonological source of confusion. For example, Schertz (2013) showed that when an initial stop's voicing was misunderstood by an interlocutor, speakers hyperarticulated VOT to minimize perceptual confusability. In this dissertation, we aimed to explore this in terms of Norwegian vowel quantity, not only looking at what adjustments speakers made but also whether these adjustments corroborated claims that adaptations can be segmentally-targeted specifically to the phonological source of confusion.

In Experiment 2, three of the four acoustic cues under investigation (i.e., vowel duration, vowel quality differences, and postvocalic consonant duration) were found to be enhanced in some way during clear speech productions: (1) vowel duration, (2) vowel quality, and (3) postvocalic consonant duration. The fourth cue of spectral movement in long mid vowels was not enhanced. While the temporal cues were enhanced across the vowel set, the acoustic distances between long and short /ɑ/ were not larger in clear speech, indicating that speakers do not enhance the quality difference between long and short /ɑ/. As Adaptive Speaker Frameworks hypothesize that speakers make local, contrast-driven adjustments to their speech, the lack of enhancement of vowel quality for /ɑ/ could raise questions about the status of vowel quality in the contrast for this vowel. Perhaps, the differences in vowel quality, while present in regular speech, are not as deeply rooted in the difference between /ɑ/ and /ɑː/ as for the other vowels. The pattern found in Experiment 3 where listeners do not use vowel quality in the perception of long and short /ɑ/ would further support this. If it's the case that vowel quality differences for long and short /ɑ/ are more a consequence of articulation than part of the contrast, the lack of enhancement and use in perception would make sense. However, a definitive explanation for this pattern requires further investigation. The data presented here offers a starting point for future research.

In addition to whether or not cues were enhanced, it is important to consider how speakers adjust them to achieve these enhancements. A documented global feature of clear speech as a stable mode of speaking is overall slower speaking rates and longer segment durations (Picheny et al., 1986). Thus, in investigating the enhancement of temporal cues in Experiment 2, the ratio of the long segment to the short segment was

used to ensure that differences in duration were not simply attributable to this global lengthening in clear speech. While there was indeed an increase in the long-to-short ratio of vowel duration, we did not see global lengthening of both long and short vowels in clear speech. In fact, short vowels were actually produced with a shorter duration in clear speech than in regular speech. This detail is important for a number of reasons. First, this goes against the notion that segments are necessarily globally lengthened in clear speech, as we see a clear example here of a segment being shortened. Second, the shortening of the short vowel (alongside the lengthening of the long vowel) is an example of a local and contrast-focused adjustment made on the part of the speaker.

A similar pattern emerges with the postvocalic consonant durations as well–while consonants after short vowels were produced longer overall, those consonants after long vowels were in fact produced with a shorter duration in clear speech. As with the pattern seen with vowel duration, this trend provides another example of speakers' enhancement strategies going against what have been deemed global characteristics of clear speech and aligning with what is proposed by Adaptive Speaker Frameworks.

In sum, this data has shown that speakers consistently perform specific and local manipulations of both temporal and spectral characteristics on Norwegian long and short vowels. This supports targeted adaptation accounts (Schertz, 2013; Buz et al., 2016, in which speakers adapt their productions when feedback from their interlocutors suggests that previous productions were perceptually confusable. Furthermore, these data support the claim that hyperarticulation can be a "targeted and flexible adaptation rather than a generalized and stable mode of speaking" (Stent et al., 2008, p. 163).

5. CUES IN PERCEPTION

When faced with phonological contrasts that are often signaled by multiple acoustic cues, it is important that listeners are able to handle these cues in perception to successfully identify phonemes. The ability of listeners to assess the informativity of an acoustic cue and adjust how heavily it is relied upon in perception is referred to as perceptual cue weighting (Holt & Lotto, 2006; Francis et al., 2008). In Experiment 3, how listeners handle the acoustic correlates of vowel quantity observed and enhanced in production was tested, specifically looking at which cues they used and how heavily they were weighted. Furthermore, it was of interest to see whether there would be vowel-specific patterns similar to what was seen in Experiments 1 and 2.

Two main patterns from the perception data in Experiment 3 are to be discussed cue weighting and ordering as it pertains to Norwegian. The first pattern is that while vowel quality was shown to be reliably used by listeners for most vowels, it was not for low vowel /ɑ/. Differences in whether or not a cue is used for all phonemes in a set for a contrast is not uncommon. Recall the case of high front vowels in Czech (Podlipsky et al., 2009) and Hungarian (Mády & Reichel, 2007), where vowel quality was used only for the perception of long and short high front vowels, but not other vowels in their experimental set. Variation in patterns seen in other languages supports claims made by previous literature that cue use and weighting is a process that is language-, dialect-, and perhaps even individual-specific (Holt & Lotto, 2006; Clayards, 2018).

The second pattern is the fact that of the vowels for which vowel quality was utilized by listeners in the categorization of long and short vowels, there were two vowels for

which listeners weighted vowel quality more heavily than vowel duration: /i/ and /e/. This pattern brings up an important point: not only can we see vowel-specific patterns in whether or not a listener utilizes a cue, but also in the weighting and relative cue ordering of the acoustic cues they use. This pattern of cues being more heavily weighted for some phonemes than for others has been attested in other languages and in other contrasts. For example, Kapnoula et al., 2017 demonstrated that the use of f0 in voicing distinction might be greater for /p-b/ distinctions than for /d-t/ (but they do not offer a concise theoretical explanation for this).

This pattern in Norwegian also brings up questions about what causes vowel quality to be weighted more heavily by listeners for these two. Returning to the explanations put forward by Holt and Lotto (2006), there might be distributional factors that cause vowel quality to be more salient than duration for listeners for these two vowels. Specifically, Holt and Lotto (2006) point to how distinctive two phonemic categories are along a given acoustic dimension as a factor in whether or not listeners will weight that cue more heavily. In Experiment 1, we saw that both /i/ and /e/ are produced with a relatively larger qualitative difference (as evidenced by the Euclidean distance) between long and short vowels. Thus, if listeners routinely hear these two vowels produced with a reliably larger qualitative difference than for other vowels, the heightened distinctiveness of vowel quality for these two vowels could lead to listeners finding this acoustic dimension more informative and relying on it more than for other vowels. The direct link between increased distinctiveness as observed in production data and the impact on listener perceptual strategies needs to be investigated further to come to a definite conclusion, especially in terms of exploring what came first. Did the increased distinctiveness in

production precede the heightened sensitivity in perception or vice versa? And what exactly has lead to speakers producing a larger difference for some vowels but not others? Nonetheless, the data here provides a suggestion that this connection is there and a motivation for future investigations.

The data from Experiment 3 are a piece in a growing body of literature aimed at understanding how listeners handle multiple acoustic cues signally a contrast, and whether these perceptual patterns are universal or language- and individual-specific. In sum, the data presented in Experiment 3 supports accounts that perceptual strategies used by listeners in weighting acoustic cues in the speech signal are highly complex and can be language, individual, and, in this case, phoneme-specific. It is not necessarily enough to look at individual languages when examining how phonological contrasts are perceived, there are often cases where we see phoneme-specific patterns, indicating a more adaptive and fine-tuned process of perception on the part of listeners.

6. RELATIONSHIP BETWEEN PRODUCTION AND PERCEPTION

While speech production and perception have traditionally been studied independently, understanding their relationship and where the converge and diverge helps to create a comprehensive model of representation and linguistic knowledge (Casserly & Pisoni, 2010; Schertz & Clare, 2020). Where these alignments and misalignments occur can have many implications for various areas of linguistic theory. For example, we can examine aspects of learning in both L1 and L2, looking at the development of the two language modalities as speakers acquire languages (Schertz & Clare, 2020). Another

example, we can look at languages as they are actively evolving, charting sound changes and cue shifts in real-time (e.g., cue shifting of tenseness marking in Southern Yi, Kuang & Cui, 2018).

Returning to the case of vowel quality in the perception of long and short /ɑ/, this is a place in the data where we saw a misalignment between production and perception. As seen in Experiment 1, long and short /ɑ/ are produced with quality differences that are significantly different, namely that long /ɑ/ is produced with a lower F2, indicating an articulation that is higher in the vowel space. Despite this difference in vowel quality in production, listeners did not use vowel quality when categorizing long and short /ɑ/. A number of studies have examined misalignments between speech production and perception, and we can consider what the misalignment shown here could mean within the context of these previous studies. In one study, Kuang and Cui (2018) examined the ongoing change in tense vs. lax register in Southern Yi; specifically, how vowel quality is overtaking phonation as the primary cue. They describe three possibilities for the time course of a cue shift: (1) shifts in production and perception simultaneously, (2) listeners shift in perception first, then mirror this in production, and (3) change occurs in production first and listeners subsequently are attuned to it in perception. Two main findings came from this study. The first is that this cue shift appeared to be more advanced in perception than production and the second is that the cue shift appeared to be less advanced in non-high vowels. This second point demonstrated that cue shifting does not need to occur uniformly across a sound system but can start in a subset of sounds before being initiated in others.

Approaching the current pattern in Norwegian with a similar angle, we can begin to speculate the reason for both this misalignment and also the fact that the misalignment is only found in one vowel pair. For example, is it possible that there is an ongoing, complete cue shifting occurring in Norwegian? Might the integration of vowel quality into the vowel quantity contrast in Norwegian be incomplete? If we consider the patterns seen in production and perception, all six vowel pairs are produced with distinct vowel qualities yet vowel quality is only used in perception for five of the six vowel pairs. This would suggest that the integration of vowel quality into the quantity contrast is more advanced in production than in perception, possibility (3) described by Kuang and Cui (2018). Furthermore, the degree to which listeners utilized vowel quality in their perception of long and short vowels varied across the vowel space: it would appear that the integration of vowel quality into the perception of vowel quantity is more advanced in non-low vowels and especially advanced for /i/ and /e/, where it is weighted more heavily than duration.

Taken together with previous literature, the data from the current dissertation demonstrates that while there is a relationship between production and perception, this relationship is complex. Furthermore, the data here show that where production and perception converge and diverge. From here we can begin to explore the status of this contrast in Norwegian and whether or not there could be a change occurring, such as what we saw in Czech (Podlipský et al., 2009) and Hungarian (Mády & Reichel, 2007). In these accounts of vowel quantity in Czech and Hungarian, the researchers point to the heavy use of vowel quality in perception as an initial sign of an incipient change in the languages, whereby a quantitative distinction between, for example, /i:/-/ɪ/ could become a qualitative distinction instead. While not showing the degradation of temporal cues in

production that is described in Czech (Podlipský et al., 2009), Norwegian listeners are using vowel quality for most vowel pairs, weighting it more heavily than vowel duration in some cases. The notion that this could occur in Norwegian is interesting to entertain and could be a point of further research in the future.

7. LIMITATIONS AND FUTURE DIRECTIONS

As with any study, this dissertation comes with a number of limitations. For example, the scope of the word types used throughout the experiments was relatively limited: in all studies, only monosyllabic CVC words were used. To further understand the acoustic realization of Norwegian vowel quantity in production and how cues are used in perception, it would be beneficial to investigate a variety of word and syllable types. Furthermore, the inclusion of polysyllabic words would enable us to look more closely at how other suprasegmental factors influence the production and perception of long and short vowels (e.g., stress or tone).

Additionally, the link between production and perception was only able to be investigated on a group-level because participants were not matched between Experiments 1, 2, and 3 (i.e., same people in each study). Of course, this group-level information is useful for understanding the link between production and perception. Yet, the ability to explore individuals would enable us to look more directly at how one's productions line up with their perceptual strategies. We could begin to answer questions like: do individuals who produce long and short vowels with a larger acoustic distance between them in the F1/F2 plane weight vowel quality more heavily in production? Data

such as this would allow us to more deeply explore both the link between production and perception and individual variation, expanding the existing literature on both of these topics.

This dissertation also serves as a foundation for future research. For example, further exploring the degree to which Norwegian listeners dynamically adjust their perceptual strategies would be an interesting future avenue. The tendency of listeners to up- or downweight cues based in perception based upon their informativity of category membership has been shown in pervious literature: Kim et al (2020) demonstrated that American English listeners will downweight spectral information and upweight vowel duration when presented with productions of the American English tense-lax distinction that are not canonical. Taking this study and applying it to Norwegian listeners would be informative in a few different areas. First, most cue weighting and learning studies relating to vowel duration and quality have been done contrasts, such as tense-lax in Kim et al. (2020), where vowel quality is the primary cue. However, such studies relating to participant reweighting of cues for long-short vowel contrasts, in which vowel duration is the primary cue, have not yet been done. Exploring this in Norwegian would add another dimension to our understanding of how these cues interact and are reweighted in a wider range of contrast types. This research would also give further insight into how Norwegian listeners handle acoustic cues in the speech signal and how they adapt to noncanonical productions of long and short vowels. For example, learners of Norwegian might produce long and short vowels will less distinct durational differences and more distinct quality differences, especially when they are native speakers of a language like English (a language without long and short vowels where duration is used a secondary cue for other

contrasts). In a situation like this, we would predict that listeners would upweight vowel quality and we would perhaps see perceptual patterns as seen for /i/ and /e/ in Experiment 3, where vowel quality is weighted more heavily than duration.

As another interesting potential avenue, a closer look at Norwegian dialects would be beneficial. Norwegian is known for having many unique and distinct dialects and the current work did not address the possibility of dialectal differences. Further investigation with a much larger participant pool and the intentional collection of data from various dialects would be beneficial for several reasons. First, it would provide an account of the acoustic realization of long and short vowels in Norwegian that is more comprehensive and nuanced. It is already an issue that acoustic-based descriptions of vowel quantity in the language are few and far between, and this dissertation aimed to begin to fill this need. However, an acoustic account that also examines the various ways in which long and short vowels are produced and perceived in Norway's various dialects would further enrich our understanding of the phonology of the language.

REFERENCES

Abramson, A., & Ren, N. (1990). Distinctive vowel length: Duration versus spectrum in Thai. *Journal of Phonetics*, *18*, 79-92.

Aoyama, Katsura (2001). *A psycholinguistic perspective on Finnish and Japanese prosody: Perception, production and child acquisition of consonantal quantity Distinctions*. Boston: Kluwer Academic Publishers.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1-48.

Beddor, P. S., Coetzee, A. W., Styler, W., McGowan, K. B., & Boland, J. E. (2018). The time course of individuals' perception of coarticulatory information is linked to their production: Implications for sound change. *Language*, *94*(4), 931-968.

Behne, D., Moxness, B., & Nylund, A. (1996). Acoustic-phonetic evidence of vowel quantity and quality in Norwegian. *Speech, Music and Hearing, Quarterly Progress and Status Report (TMH-QPSR)*, *37*(2), 13-16.

Behne, D., Czigler, P., & Sullivan, K. P. (1997). Swedish quantity and quality: a traditional issue revisited. *Reports from the Department of Phonetics, Umeå University*, *4*, 81-83.

Behne, D., Arai, T., Czigler, P., & Sullivan, K. (1999). Vowel duration and spectra as perceptual cues to vowel quantity: A comparison of Japanese and Swedish. In *Proceedings of the 14th International Congress of Phonetic Sciences*, 857-860.

Behne, D. M., & Nylund, A. (2003). Desensitization in Norwegian vowel perception by native and American English listeners. In *Proceedings of the 15th International Congress of Phonetic Sciences*, 1557-1560.

Benguerel, A.P. & Bhatia, T. K. 1980. Hindi stop consonants: An acoustic and fiberscopic study. *Phonetica*, *37*, 134–148.

Boersma, P. & Weenink, D. (2022). Praat: doing phonetics by computer (Version 6.3.02) [Computer program].

Bond, Z. S., Moore, T. J., & Gable, B. (1989). Acoustic–phonetic characteristics of speech produced in noise and while wearing an oxygen mask. *The Journal of the Acoustical Society of America, 85*(2), 907-912.

Bradlow, A. R. (2002). Confluent talker- and listener-related forces in clear speech production. In C. Gussenhoven and N. Warner (Ed.), *Laboratory Phonology* (Vol. 7, pp. 241-273). Berlin: Mouton de Gruyter.

Buz, E., Tanenhaus, M. K., & Jaeger, T. F. (2016). Dynamically adapted context-specific hyper-articulation: Feedback from interlocutors affects speakers' subsequent pronunciations. *Journal of Memory and Language*, *89*, 68-86.

Casserly, E. D., & Pisoni, D. B. (2010). Speech perception and production. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(5), 629-647.

Chen, F. R. (1980). *Acoustic characteristics and intelligibility of clear and conversational speech at the segmental level* [Doctoral dissertation, Massachusetts Institute of Technology].

Chen, M. (1970). Vowel length variation as a function of the voicing of the consonant environment. *Phonetica*, *22*(3), 129-159.

Clayards, M. (2018). Differences in cue weights for speech perception are correlated for individuals within and across contrasts. *The Journal of the Acoustical Society of America, 144*(3), EL172-EL177.

Corretge, R. (2012). Praat Vocal Toolkit [Software package].

Cohn, M., Segedin, B. F., & Zellou, G. (2022). Acoustic-phonetic properties of Siri-and human-directed speech. *Journal of Phonetics*, *90*, 101-123.

Davis, S. (2011). Quantity. *The handbook of phonological theory*, 103-140.

Elert, C.C. (1964). *Phonological studies of quantity in Swedish*. Stockholm: Almqvist & Wiksell.

Eliasson, S. (1985). Stress alternations and vowel length: new evidence for an underlying nine vowel system in Swedish. *Nordic Journal of Linguistics*, *8*, 101–129.

Escudero, P., Benders, T., & Lipski, S. C. (2009). Native, non-native and L2 perceptual cue weighting for Dutch vowels: The case of Dutch, German, and Spanish listeners. *Journal of Phonetics*, *37*(4), 452-465.

Fintoft, K. (1961). The duration of some Norwegian speech sounds. *Phonetica*, *7*(1), 19-39.

Fretheim, T. (1969). Norwegian stress and quantity reconsidered. *Norsk Tidsskrift for Sprogvidenskap*, *23*, 76–96.

Garnes, S. (1976). *Quantity in Icelandic: Production and perception*. Hamburg: Helmut Buske Verlag.

Grenon, I., Kubota, M., & Sheppard, C. (2019). The creation of a new vowel category by adult learners after adaptive phonetic training. *Journal of Phonetics*, *72*, 17-34.

Ham, W.H. (2001). *Phonetic and Phonological Aspects of Geminate Timing*, New York: Routledge.

Hillenbrand, J. M., Clark, M. J., & Houde, R. A. (2000). Some effects of duration on vowel recognition. *Journal of the Acoustical Society of America*, *108*, 3013–3022.

Holt, L. L., & Lotto, A. J. (2006). Cue weighting in auditory categorization: Implications for first and second language acquisition. *The Journal of the Acoustical Society of America*, *119*(5), 3059-3071.

Idemaru, K., & Holt, L. L. (2011). Word recognition reflects dimension-based statistical learning. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(6), 1939-1956.

Jahr, E.H., & Lorentz, O. (1983). *Prosodi/Prosody*. Oslo: Novus.

Kapnoula, E. C., Winn, M. B., Kong, E. J., Edwards, J., & McMurray, B. (2017). Evaluating the sources and functions of gradiency in phoneme categorization: An individual differences approach. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(9), 1594-1611.

Keyser, S. J., & Stevens, K. N. (2006). Enhancement and Overlap in the Speech Chain. *Language*, *82*(1), 33-63.

Kim, D., Clayards, M., & Kong, E. J. (2020). Individual differences in perceptual adaptation to unfamiliar phonetic categories. *Journal of Phonetics*, *81*, 100984.

Kingston, J., & Diehl, R. L. (1994). Phonetic knowledge. *Language*, *70*(3), 419-454.

Kinoshita, K., Behne, D. M., & Arai, T. (2002). Duration and F0 as perceptual cues to Japanese vowel quantity. *Perception*, *145*(4), 757-760.

Krause, J., & Braida, L. (2004). Acoustic properties of naturally produced clear speech at normal speaking rates. *Journal of the Acoustical Society of America*, *115*(1), 362-378.

Kristoffersen, G. (2000). *The Phonology of Norwegian*. Oxford: Oxford University Press.

Kristoffersen, G. (2011). Quantity in Old Norse and modern peninsular North Germanic. *The Journal of Comparative Germanic Linguistics*, *14*, 47-80.

Kuang, J., & Cui, A. (2018). Relative cue weighting in production and perception of an ongoing sound change in Southern Yi. *Journal of Phonetics*, *71*, 194-214.

Kvifte, B., & Gude-Husken, V. (2005). *Übungsbuch zur norwegischen Grammatik: mit einem Schlüssel zu den Übungen*. Wilhelmsfeld.

Lehiste, I. (1976). Influence of fundamental frequency patterns on the perception of duration. *Journal of Phonetics*, *4*, 113-117.

Lehnert-LeHouillier, H. (2010). A cross-linguistic investigation of cues to vowel length perception. *Journal of Phonetics*, *38(*3), 472-482.

Lenth, R., Herve, M., Love, J., Riebl, H., & Singman, H. (2021). Package ‘emmeans’ [Software package].

Liberman, A.M., Cooper, F.S., Shankwiler, D.P., & Studdert-Kennedy, M. (1967). Perception of the Speech Code. *Psychological Review*, *74*, 431-461.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. *Speech Production and Speech Modelling*, 403-439.

Lippus, P. (2011). *The acoustic features and perception of the Estonian quantity system* [Doctoral dissertation, Tartu University].

Lippus, P., Asu, E. L., Teras, P., & Tuisk, T. (2013). Quantity-related variation of duration, pitch and vowel quality in spontaneous Estonian. *Journal of Phonetics*, *41*(1), 17-28.

Lisker, L., and Abramson, A. S. (1964). A cross-linguistic study of voicing in initial stops: Acoustical measurements. *Word*, *20*, 384–422.

Liu, R., & Holt, L. L. (2015). Dimension-based statistical learning of vowels. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(6), 1783.

Lutfi, R. A., & Doherty, K. D. (1994). The role of component-relative entropy in the discrimination of simultaneous tone complexes. *The Journal of the Acoustical Society of America*, *95*(5), 2963-2963.

Maddieson, I. (1984). *Patterns of sounds*. Cambridge: Cambridge University Press.

Mády, K., Reichel, U.D. (2007). Quantity distinction in the Hungarian vowel system–just theory or also reality? In *Proceedings of the 16th International Congress of Phonetic Science*, 1053-1056.

Mitterer, H., & Ernestus, M. (2008). The link between speech perception and production is phonological and abstract: Evidence from the shadowing task. *Cognition*, *109*(1), 168-173.

Moon, S.J., & Lindblom, B. (1994). Interaction between duration, context, and speaking style in English stressed vowels. *Journal of the Acoustical Society of America*, *96*(1), 40-55.

Morrison, G. S. (2005). An appropriate metric for cue weighting in L2 speech perception: Response to Escudero and Boersma (2004). *Studies in Second Language Acquisition*, *27*(4), 597-606.

Morrison, G. S. (2007). Logistic regression modelling for first and second language perception data. *Amsterdam Studies in the Theory and History of Linguistic Science Series*, *4,* 219-282

Nearey, T. M. (1977). *Phonetic feature systems for vowels*. [Doctoral dissertation, University of Alberta].

Nearey, T. M. (1997). Speech perception as pattern recognition. *The Journal of the Acoustical Society of America, 101*(6), 3241-3254.

Newman, R. S. (2003). Using links between speech perception and speech production to evaluate different acoustic metrics: A preliminary report. *The Journal of the Acoustical Society of America*, *113*(5), 2850-2860.

Nylund, A., & Behne, D. M. (1996). Acoustic characteristics of perceived vowel quantity and quality in English and Norwegian vowels. *The Journal of the Acoustical Society of America*, *100*(4), 2686-2687.

Ohala, J. J. (1994). Clear speech does not exaggerate phonemic contrast. *The Journal of the Acoustical Society of America*, *96*(5), 3227-3227.

Oviatt, S., Bernard, J., & Levow, G. A. (1998). Linguistic adaptations during spoken and multimodal error resolution. *Language and speech*, *41(*3-4), 419-442.

Page, B. (2020). Quantity in Germanic Languages. In M. Putnam & B. Page (Eds.), *The Cambridge Handbook of Germanic Linguistics* (pp. 97-118). Cambridge: Cambridge University Press.

Payne, E., Post, B., Garmann, N.G., & Simonsen, H.G. (2017). The acquisition of long consonants in Norwegian. In H. Kubozono (Ed.), *The Phonetics and Phonology of Geminates* (Vol. 2, pp. 130-162). Oxford: Oxford University Press.

Peterson, G. E., & Lehiste, I. (1960). Duration of syllable nuclei in English. *The Journal of the Acoustical Society of America*, *32*(6), 693-703.

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1986). Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech. *Journal of Speech, Language, and Hearing Research, 29*(4), 434-446.

Pind, J. (1996). Spectral factors in the perception of vowel quantity in Icelandic. *Scandinavian Journal of Psychology*, *37*(2), 121-131.

Pisoni, D. B. (1976). Fundamental frequency and perceived vowel duration. *The Journal of the Acoustical Society of America*, *59(*S1), S39-S39.

Podlipský, V. J., Skarnitzl, R., & Volín, J. (2009). High front vowels in Czech: A contrast in quantity or quality?. In *Tenth Annual Conference of the International Speech Communication Association*.

Ren, Ruqin (2018, February 16). Testing the equality of coefficients in the same regression model (comparison of two or more coefficients). https://ruqinren.wordpress.com/2018/02/16/testing-the-equality-of-coefficients-in-the-same-regression-model/.

Riad, Tomas (1992). *Structures in Germanic Prosody. A diachronic study with special reference to the Nordic languages*. [Doctoral dissertation, Stockholm University].

Rosen, S. M. (1977). Fundamental frequency patterns and the long-short vowel distinction in Swedish. *Speech Transmission Laboratory, Quarterly Progress and Status Report*, *1*, 31-37.

Rostolland, D. (1982). Acoustic features of shouted voice. *Acta Acustica United with Acustica*, *50*(2), 118-125.

Scarborough, R., & Zellou, G. (2013). Clarity in communication: "Clear" speech authenticity and lexical neighborhood density effects in speech production and perception. *The Journal of the Acoustical Society of America*, *134*(5), 3793-3807.

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

Schertz, J., Cho, T., Lotto, A., & Warner, N. (2015). Individual differences in phonetic cue use in production and perception of a non-native sound contrast. *Journal of Phonetics*, *52*, 183-204.

Schertz, J., Carbonell, K., & Lotto, A. J. (2020). Language specificity in phonetic cue weighting: Monolingual and bilingual perception of the stop voicing contrast in English and Spanish. *Phonetica*, *77(*3), 186-208.

Schertz, J., & Clare, E. J. (2020). Phonetic cue weighting in perception and production. *Wiley Interdisciplinary Reviews: Cognitive Science*, *11*(2), e1521.

Seifert, L. W. (1963). Stress Accent in Dane County Kölsch. *Monatshefte*, 195-202.

Sendelmeier, W. (1981). Der Einfluss von Qualität und Quantität auf die Perzeption betonter Vokale des Deutschen. *Phonetica*, *38*, 291-308.

Smiljanic, R., & Bradlow, A. R. (2008). Stability of temporal contrasts across speaking styles in English and Croatian. *Journal of Phonetics*, *36*(1), 91-113.

Solé, M. J., & Ohala, J. J. (2010). What is and what is not under the control of the speaker: Intrinsic vowel duration. *Papers in Laboratory Phonology*, *10*, 607-655.

Stausland Johnsen, S. (2019, February 22). *Does secondary stress exist in Norwegian?* [Paper presentation]. Phonology in the Nordic Countries 2019, Edinburgh, Scotland.

Stent, A. J., Huffman, M. K., & Brennan, S. E. (2008). Adapting speaking after evidence of misrecognition: Local and global hyperarticulation. *Speech Communication*, *50*(3), 163-178.

Stevens, K. N., & Keyser, S. J. (1989). Primary features and their enhancement in consonants. *Language*, *65*(1), 81-106.

Thomas, E. R., & Kendall, T. (2007). NORM: The vowel normalization and plotting suite. http://ncslaap.lib.ncsu.edu/tools/norm/index.php.

Traunmüller, H.,& Krull, D. (2003). The effect of local speaking rate on the perception of quantity in Estonian. *Phonetica*, *60*, 187–207.

van Dommelen, W.A. (1993). Does dynamic F0 increase perceived duration? New light on an old issue. *Journal of Phonetics*, *21*(4), 367-386.

van Dommelen, W. A. (1999). Auditory accounts of temporal factors in the perception of Norwegian disyllables and speech analogs. *Journal of Phonetics, 27*(1), 107-123.

Vanvik, A. (1972). A phonetic-phonemic analysis of Standard Eastern Norwegian. Part I. *Norwegian Journal of Linguistics*, *26*(2), 119-164.

Wang, W. S. Y., Lehiste, I., Chuang, C. K., & Darnovsky, N. (1976). Perception of vowel duration. *The Journal of the Acoustical Society of America*, *60*(S1), S92-S92.

Wheeler, A. P. (2016, October 19). Testing the equality of two regression coefficients. https://andrewpwheeler.com/2016/10/19/testing-the-equality-of-two-regression-coefficients/.

Zellou, G. (2017). Individual differences in the production of nasal coarticulation and perceptual compensation. *Journal of Phonetics, 61*, 13-29.