NORTHWESTERN UNIVERSITY

Phonotactic Similarity as a Predictor of Perceived Vowel Similarity

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Linguistics

By

Züheyra Tokaç

EVANSTON, ILLINOIS

December, 2025

Abstract

From an exemplar models perspective, linguistic category representations emerge as generalizations over exemplar clouds as a result of similarity-based coactivation of exemplars. This proposal can be applied to vowel harmony whereby harmony classes are construed as mental representations that emerge from vowel cooccurrence patterns in the lexicon. Under this view, vowel cooccurrence is part of the information encoded when storing exemplars, and a harmony class is an exemplar cloud consisting of vowels coactivated by cooccurring in a lexical exemplar. A prediction of exemplar models is that perceived similarity of exemplars will be higher within a category than across categories. In the context of harmony classes, the prediction is that perceived vowel similarity will be higher within a harmony class than across harmony classes. In this work, I propose a link between the similarity-based coactivation of cooccurring vowels and perceived vowel similarity.

Based on this proposal, I ask whether vowel cooccurrence patterns predict perceived vowel similarity in a vowel harmony language. I examine perceived vowel similarity in Turkish, a language noted for its highly systematic vowel harmony. However, Turkish vowel harmony incompletely applies in the lexicon: words in which disharmonic vowels cooccur are also attested. Moreover, vowel pairs within a harmony class as well as disharmonic vowel pairs vary in their likelihood of cooccurrence in the lexicon. Based on this variability, I hypothesize that higher likelihood of vowel cooccurrence predicts higher perceived vowel similarity. To this end, I explore the phonetic predictors of vowel similarity in Turkish in a vowel production experiment, and examine perceived vowel similarity in vowel identification and discrimination experiments.

In the Turkish vowel production experiment (Chapter 2), I find that Turkish vowels are continuously distributed along the F2 dimension, suggesting weak phonetic contrasts, and that vowels participate in phonological processes as individual vowels rather than as harmony classes.

I also find that peripheral vowels with relatively more extreme formant values exhibit phonetic variability to a greater degree across various contexts. I argue that these patterns of variation are motivated by contrast enhancement and contrast preservation in the Turkish vowel space.

In the vowel discrimination experiments (Chapter 3), I show that phonological predictors of vowel similarity contribute to perceived vowel similarity beyond phonetic predictors. Results also suggest that higher similarity in vowel features predicts higher perceived vowel similarity in Turkish but not in English. I argue that vowel features are more salient in Turkish vowel perception as a result of experience with Turkish vowel harmony.

To test my main hypothesis that higher likelihood of vowel similarity predicts higher perceived vowel similarity, I analyze perceptual confusions in vowel identification experiments (Chapter 4). I find evidence supporting this hypothesis in English vowel identification. However, contrary to my hypothesis, higher likelihood of vowel cooccurrence predicts *lower* perceived vowel similarity in Turkish. I argue that this unexpected finding suggests that the likelihood of vowel cooccurrence serves a function of enhancing perceptual contrasts rather than enhancing perceptual similarity between vowels in Turkish.

Results from post-hoc analyses suggest that higher likelihood of vowel cooccurrence predicts higher perceived vowel similarity as hypothesized for peripheral Turkish vowels, although the opposite pattern is observed for nonperipheral Turkish vowels. In other words, the hypothesized relationship does not generalize to all Turkish vowels. I discuss this finding in relation to the claim that Turkish vowel harmony is perceptually motivated to facilitate the accurate perception of nonperipheral vowels (Suomi, 1983). I argue that Turkish vowel harmony might contextually enhance the phonetically and perceptually weak contrasts between nonperipheral and peripheral vowels in Turkish.

Acknowledgements

First and foremost, I would like to thank my advisor Jennifer Cole and the members of my committee, Ann Bradlow and Matt Goldrick. I am forever inspired by their scientific rigor and trailblazing research. Jennifer helped me grow as a scientist and guided and supported me throughout this journey in many ways beyond being a wonderful advisor academically.

It was a privilege being part of the Linguistics community at Northwestern University. First, I must thank Chun Lian Chang, for we can never thank him enough. I thank Airica Thomas and Michael Dover for their moral support. I believe we would make a great cohort in another, more fortunate lifetime. Past and present members of ProSD Lab have helped me develop research questions, methods, arguments. I am very thankful to each and every one of them, and to Kate Sandberg, Shawn Foster, and Dan Turner for their friendship. Special thanks to Chantal De Leon, Raef Khan, Ziyun Chew, and, once more, Airica Thomas. You became more than friends to me, and I already miss you so much.

I found home at Northwestern with Mim Khadija Ahmed, Mame Coumba Ka, Rama Darayyad, and Mairi Glynn. I cannot begin to express how much you mean to me and I thank you with my whole heart. I am beyond thankful to Beria Tanrıöver. You were always there for me and always will be. I thank Hilal Dağlar and Ayşegül Çaşkurlu Öz for endless hours of great conversation.

Women’s friendships are the most special thing in this life, and I am blessed with many. I thank the friends that I made in high school for they are the friends who made me—Dilara Zengince, Nurcihan Ekleme, Betül Balaban Baran, Feyzanur Kurtaran Kaya, Elif Nur Özervarlı, Aslı Berre Öztürk, Müberra Özkan, Meryem Seda Güner, Senanur Kılınç, among others. And I am thankful to my friends from college—Elifnur Özden, Saliha Erman, Seda Karaköse Akbıyık, Banu Gül, Hilal Çakır, Esra Asrak Büyük, Merve Funda Cankurtaran, Merve Bilgili, and Didem

Kaya Bayram. A friendship like ours is truly a blessing. A heartfelt thank you to my dear friend Şuheda Ece Sarı and especially to Serra Çınar forever and ever.

I thank my family from the bottom of my heart. I thank my parents for everything and my sisters for being the best friends one could ask for. It is almost time we stopped collecting graduate degrees like it is a family tradition we cannot seem to break. I am very, very thankful for my niece, our most precious.

I thank Keyman Modern Turkish Studies Program at Northwestern University for funding that supported my research, Northwestern University Turkish Students Association and the Turkish community in Chicago for volunteering their time to participate in my experiments, my participants in Istanbul and elsewhere, and every single person who recorded stimuli for me. Lastly, I thank Northwestern University Graduate Workers for all the incredibly important work they do.

ب

Table of Contents

List of Tables

Table of Figures

CHAPTER 1

INTRODUCTION

Exemplar models provide a unique perspective of linguistic representations that emerge from the rich information that is preserved in encoding the signal. The classic example for describing the premise of speech perception under exemplar models is one of vowel identification (e.g., Johnson, 1997; Pierrehumbert, 2001). The exemplar space for vowels consists of the memory traces of each vowel token encountered. Each token is parameterized along phonetic dimensions such as vowel formants, F0, or duration. Stored tokens are organized in an exemplar space with exemplars with similar phonetic parameter values situated closer. When a new vowel token is encountered, it is mapped in this exemplar space, which activates the preexisting vowel exemplars within a distance as a function of their similarity to the new token. The new token is identified probabilistically as a member of the vowel category with the most exemplars coactivated. The new token is then labeled and added to the cloud of exemplars for that vowel category. From this perspective, a vowel category is an emergent construct defined as a set of exemplars that are phonetically similar and have a high likelihood of coactivation.

Similar to the emergent vowel phoneme representations, exemplars of words are stored with phonetic detail which lead to emergent lexical representations. Other linguistic representations might emerge from the similarity-based coactivation of exemplars along other dimensions such as social, phonological, or lexical. In this dissertation, I focus on vowel harmony classes as emergent linguistic representations from vowel cooccurrence patterns in the lexicon and ask what perceptual consequences might follow from such a conceptualization. I argue that exemplars that are similar are coactivated, which is in turn reflected in higher perceived similarity. In the case of harmony classes, I hypothesize that higher likelihood of vowel cooccurrence in the lexicon leads to higher perceived vowel similarity.

I examine vowel similarity as perceived by native listeners in Turkish, a language with vowel harmony. In a series of vowel discrimination and vowel identification experiments, I show that perceived vowel similarity is complex and multilayered, combining phonetic and phonological measures of vowel similarity as well as likelihood of vowel cooccurrence in the lexicon. However, contrary to my hypothesis, I find that likelihood of vowel cooccurrence is not unilaterally positively related to perceived vowel similarity. I propose an alternative explanation that likelihood of vowel cooccurrence interacts with the perceptual confusability of vowels as predicted by vowel phonetics. I interpret this interaction to suggests that contrast preservation as a perceptual goal interferes with enhanced perceptual similarity of vowels in harmony class formation. In this chapter, I introduce my hypothesis in more detail and preview the discussion to follow from the unexpected results.

1.1. Harmony classes as emergent representations

The proposed mechanism of category formation in exemplar models is similarity-based coactivation. This mechanism can be applied to phonemic as well as other levels of linguistic representation. For instance, a token of the word '*bay*' will activate the stored exemplars of '*bay*' but also might activate stored exemplars of '*bell*', '*bit*', and other words that begin with [b]. The set of exemplars that are coactivated in such instances define an emergent construct of syllable- and word-initial [b]. More complex patterns can also be described in terms of similarity-based coactivation of exemplars. Consider a vowel harmony language in which a word such as /ibe/ activates other words in which the vowel [i] is in the same position, and the coactivated set of words include /ibi/, /ime/ and /imi/ where the second vowel is a front vowel. On the other hand, a word such as /aba/ coactivates other words such as /abu/ and /abo/, where the second vowel is a back vowel. That is, a front vowel in the first syllable coactivates words with front vowels in the second syllable, and a back vowel in the first syllable coactivates words with vowels in the second syllable. From these coactivation patterns based on vowel cooccurrence similarity, classes of front-harmonic and back-harmonic words might emerge. The vowels in words that are coactivated according to vowel cooccurrence patterns might also lead to harmony classes to emerge, such as front vowel and back vowel classes in this example. This understanding of emergent harmony classes contrasts with traditional views of vowel harmony where harmony features are presupposed (Cole, 2009).

My work asks questions relating to this proposed relationship between vowel cooccurrence patterns in the lexicon and emergent harmony classes. A prediction of exemplar models given the essential role of similarity-based coactivation in emergent category representations is that such a relationship would be reflected in perceived vowel similarity. One way in which category representations impact perceived similarity is that perceptual similarity is higher within categories than across. Discrimination of speech sounds that are equidistant phonetic steps from each other is harder when the two sounds are identified as members of the same phoneme category than when they are identified as members of two distinct phoneme categories. Moreover, within a category, discrimination sensitivity lowers as the tokens approach the best exemplar of the category (e.g., Iverson & Kuhl, 2000; Kuhl, 1991).

These findings suggest a discrepancy between phonetic and perceived distance that is mediated by coactivation. Whereas phonetic distance is encoded in the exemplar space, it does not directly translate into perceived distance. Instead, coactivation of exemplars within a category warps the exemplar space. Tokens that are coactivated to a greater degree are perceived to be more similar whereas tokens that are not coactivated and hence are members of distinct categories are perceived to be more dissimilar than their raw phonetic similarity predicts. Applied to the case of harmony classes, which consist of more than one vowel category, a similar yet potentially weaker effect might be predicted: vowels belonging to the same harmony class might be perceived to be more similar than vowels belonging to different harmony classes.

Findings from a cross-linguistic perceived vowel similarity rating study support the prediction above. Terbeek (1977) presented listeners of various native languages with nonnative vowel triplets and asked them to indicate which vowel pair they thought sounded most similar and most dissimilar. Across language groups, nonnative vowels that map onto native vowel contrasts were rated to be more dissimilar. On the other hand, vowels that map onto the same native vowel category were rated to be more similar. This finding confirms that perceived similarity is warped by category membership as suggested above. Moreover, native Turkish listeners rated pairs of front and back vowels to be more dissimilar than other language groups, suggesting that the phonetic distance between front and back vowels are perceptually enhanced for Turkish listeners. Terbeek attributes this distinct pattern observed in Turkish listeners to Turkish vowel harmony. That is, as suggested above, vowels belonging to different harmony classes were perceived to be more dissimilar than their phonetic similarity predicts as well as compared to their perceived similarity in other, non-harmony languages.

The interpretation of Terbeek's (1977) finding from the perspective of exemplar models is based on the proposal that harmony classes emerge from the vowel cooccurrence patterns in the lexicon. This proposal entails an exemplar space that is organized according to vowel cooccurrence. That is, exemplars where two vowels cooccur (e.g., /abu/, /uba/, /uma/) would be organized closely such that they are coactivated. This coactivation based on vowel cooccurrence would spread to other exemplars with partially overlapping vowel cooccurrence patterns (e.g., /aba/, /abo/, /obu/). Coactivation would be lowest for exemplars with vowel cooccurrence patterns with no overlap (e.g., /ebi/, /ime/). As a result, spreading coactivation would be greater for vowels that are more similar in their cooccurrence patterns, leading to harmony classes to emerge over these sets of exemplars that are coactivated. Lesser coactivation across harmony classes predicts lower perceived similarity as observed by Terbeek. Whereas Terbeek's analysis focuses on perceived vowel similarity compared across harmony classes, the prediction above can also be extended to any individual vowel pair as follows.

The simplified example above of vowel cooccurrence patterns in a lexicon assumes that all exemplars exhibit harmonic vowel cooccurrence patterns. In contrast, evidence suggests that Turkish vowel harmony incompletely applies in the lexicon: harmonic vowel cooccurrence patterns are observed at varying frequencies across the vowels and disharmonic vowel cooccurrence patterns are also attested at a nonnegligible rate (e.g., Kabak et al., 2008). Hence, the above example can be updated to capture the rates at which each vowel cooccurrence pattern is observed and to include spreading coactivation to disharmonic exemplars as well, such as /abi/ and /ube/. In such a case, the degree of coactivation for any vowel pair would be predicted by the likelihood of observing each of these vowel cooccurrence patterns. This second prediction leads me to hypothesize that higher likelihood of vowel cooccurrence in the lexicon would lead to higher perceived vowel similarity beyond phonetic similarity as well as beyond agreement in vowel features predict.

One might argue that this hypothesis is counterintuitive on the grounds that coactivation-led higher perceived vowel similarity might eventually lead to contrast neutralization. For instance, sustained coactivation of vowels within the front and the back harmony classes might reiteratively enhance perceived vowel similarity to the point that only two vowels remain in the language: one front vowel and one back vowel. Alternatively, higher likelihood of vowel cooccurrence might be predicted to help establish contrasts as a result of vowels within a harmony class having overlapping distributions in the lexicon. For instance, recognizing the lexical contrast between words such as /abu/ and /ubu/ requires recognizing the phonemic distinction between /a/ and /u/. To anticipate the discussion and the post-hoc analysis in Chapter 4, my findings suggest that the contribution of likelihood of vowel cooccurrence on perceived vowel similarity is relatively small compared to the influence of phonetic similarity of vowels, and that the effect of likelihood of vowel cooccurrence interacts with phonetic similarity, presumably to the goal of contrast preservation between vowels that are perceptually highly confusable.

To reiterate, the proposal that harmony classes emerge from vowel cooccurrence patterns in the lexicon entails an exemplar space organized according to cooccurrence similarity on a level of representation beyond the phonetic level. I argue that cooccurrence leads to coactivation of exemplars, which might lead to emergent harmony classes. I hypothesize that two vowels with relatively higher likelihood of cooccurrence will be coactivated more and therefore perceived to be more similar than their phonetic similarity alone would predict. I test this hypothesis in a series of experiments looking at native vowel perception in Turkish, a language with pervasive and systematic vowel harmony and yet variability in the degree to which vowel harmony applies across the lexicon. To describe this variability, in the next section, I present an overview of Turkish vowel harmony and the vowel cooccurrence patterns observed in the Turkish lexicon.

1.2. Turkish vowel harmony

Turkish has an inventory of eight vowels which are traditionally described by a combination of three features: backness (front and back), rounding (rounded and unrounded), and height (high and nonhigh). The Turkish vowel inventory is completely symmetrical with respect to these vowel features, with four front and four back vowels, four rounded and four unrounded vowels, and four high and four nonhigh vowels (see Table 1.1). This description of the Turkish vowels in terms of their phonological features does not directly follow from Turkish vowel phonetics (e.g., Kopkallı-Yavuz, 2010; see Chapter 2). Rather, it is participation of the Turkish vowels in vowel harmony that leads to this description as follows.

Table 1.1. The Turkish vowel system.

	front		back	
	unrounded	rounded	unrounded	rounded
high	i	y	ɯ	u
nonhigh	e	œ	ɑ	o

The domain of vowel harmony in Turkish is the word. In simplest terms, Turkish vowel harmony is the set of conditions that within a word, (1) vowels agree with the preceding vowel in backness and (2) high vowels agree with the preceding vowel in rounding (e.g., Kornfilt, 1997). Table 1.2 and Table 1.3 represent the two conditions, respectively, in terms of permissibility in vowel sequences based on agreement in backness and rounding. In these tables, rows represent the first vowel of the vowel sequence, or the trigger, and the columns represent the second vowel, or the target. For each vowel pair, a plus sign indicates a permissible sequence whereas a minus sign indicates an impermissible sequence.

Table 1.2. Permissible vowel sequences based on vowel backness in Turkish.

	V_2 (target)	
V_1 (trigger)	back	front
back	+	-
front	-	+

Table 1.3. Permissible vowel sequences based on vowel height and rounding in Turkish.

	V_2 (target)	
V_1 (trigger)	high rounded	high unrounded
rounded	+	-
unrounded	-	+

These conditions of agreement in backness and rounding apply to native Turkish roots as well as suffixes, suggesting that the vowel cooccurrence tendencies in roots and suffixes are a unified phenomenon (see Kabak, 2011). However, Turkish vowel harmony is manifested in productive vowel alternations only in suffixation. In contrast, in root internal harmony, it is manifested as vowel cooccurrence tendencies in the lexicon. The Turkish words in (1) demonstrate cases where these conditions are met root-internally and illustrate the backness- and rounding-harmonizing vowel alternations in the suffix. A subset of the roots in (1) are repeated in (2) with a suffix that alternates in backness only.

(1)	/ɑslɑn-ɯ/	'lion-ACC'	/ejer-i/	'saddle-ACC'
	/bɑkɯr-ɯ/	'copper-ACC'	/deniz-i/	'sea-ACC'
	/pɯnɑr-ɯ/	'spring-ACC'	/t͡ʃimen-i/	'grass-ACC'
	/ɯlgɯn-ɯ/	'tamarisk-ACC'	/titiz-i/	'neat-ACC'
	/ozɑn-ɯ/	'bard-ACC'	/œzen-i/	'diligence-ACC'
	/ojun-u/	'game-ACC'	/gœnyl-y/	'heart-ACC'
	/juvɑr-ɯ/	'sphere-ACC'	/gyzel-i/	'beautiful-ACC'
	/t͡ʃukur-u/	'pit-ACC'	/yzym-y/	'grape-ACC'
(2)	/ɑslɑn-ɑ/	'lion-DAT'	/ejer-e/	'saddle-DAT'
	/bɑkɯr-ɑ/	'copper-DAT'	/deniz-e/	'sea-DAT'
	/t͡ʃukur-ɑ/	'pit-DAT'	/yzym-e/	'grape-DAT'

The roots in (1) and (2) also demonstrate two other tendencies regarding vowel distribution in Turkish words (e.g., Erguvanlı-Taylan, 2015; Zimmer, 1969). First, all 8 Turkish vowels are observed in word-initial syllables. Second, nonhigh rounded vowels /o œ/ are not observed in non-initial syllables. Similarly, there are no suffixes with an /o-œ/ alternation. The combined set of vowel cooccurrence constraints and tendencies summarized above can also be stated in terms of whether each individual vowel sequence is permissible or not as in Table 1.4. As in Table 1.2 and Table 1.3, a plus sign indicates a permissible sequence whereas a minus sign indicates an impermissible sequence. Table 1.4 differs in that this table represents all possible sequences of vowels as defined by the three vowel features. Notice that the top right and the bottom left quadrants populated by front-back or back-front vowel sequences consist of minus signs only. The plus signs are confined to the top left and bottom right quadrants where vowels agree in backness, although these cells also reflect agreement in rounding where relevant.

Table 1.4. Permissible vowel sequences in Turkish (partially reproduced from Mayer et al., 2010).

	V_2 (target)							
V_1 (trigger)	ɑ	ɯ	o	u	e	i	œ	y
ɑ	+	+	-	-	-	-	-	-
ɯ	+	+	-	-	-	-	-	-
o	+	-	-	+	-	-	-	-
u	+	-	-	+	-	-	-	-
e	-	-	-	-	+	+	-	-
i	-	-	-	-	+	+	-	-
œ	-	-	-	-	+	-	-	+
y	-	-	-	-	+	-	-	+

What is not captured by the description of Turkish vowel harmony above is that disharmony is observed in some Turkish roots as well as suffixes. In fact, all possible vowel sequences listed in Table 1.4 are observed in the Turkish lexicon, although harmonic sequences are generally more prevalent. For instance, previous research finds that 29% of Turkish roots in a corpus are backness disharmonic in their first to syllables and that some disharmonic vowel sequences are more frequent than some harmonic vowel sequences (Kabak et al., 2008; Kabak & Weber, 2013). Similarly, Güngör (2003) finds that 41.2% of the stems in a corpus made of three Turkish dictionaries are backness disharmonic. In other words, Turkish vowel harmony incompletely applies across the lexicon despite its discrete description as in Tables 1.2-1.4. This seeming discrepancy arises from viewing harmony as the input and the observed cooccurrence patterns as the output. Alternatively, the observed cooccurrence patterns can be viewed as the probabilistic input from which harmony is derived as the output (cf. Caplan & Kodner, 2018; Baker, 2009; Cole, 2009).

Cole (2009) proposes that vowel cooccurrence patterns in the Turkish lexicon can lead to the emergence of harmony classes from a measure as simple as bigram transition probabilities of individual vowels. Others have used similar measures to show that likelihood of vowel cooccurrence varies across vowel sequences and yet, on the whole, yields patterns that are aligned with the discrete descriptions of vowel harmony in Turkish (Baker, 2009; Caplan & Kodner, 2018; Kabak & Weber, 2013; Mayer et al., 2010). These measures of association strength indicate whether compared to what would be expected under independent distribution of vowels, vowel sequences have higher likelihood of being observed or not.

For instance, in their models, Caplan and Kodner (2018) divide the bigram probability of /$V_1C^+V_2$/ by the unigram probability of /V_1/ to measure the likelihood of a vowel sequence while also controlling for individual vowel frequency. Figure 1.1 represents the likelihood of vowel sequences in Turkish in a heatmap. Green cells indicate that the vowel sequence is observed with higher likelihood than the expected probability. Red cells indicate that the vowel sequence is observed with lower likelihood than the expected probability. Lastly, darker shades represent a stronger divergence from expected probability under independent distribution of vowels.

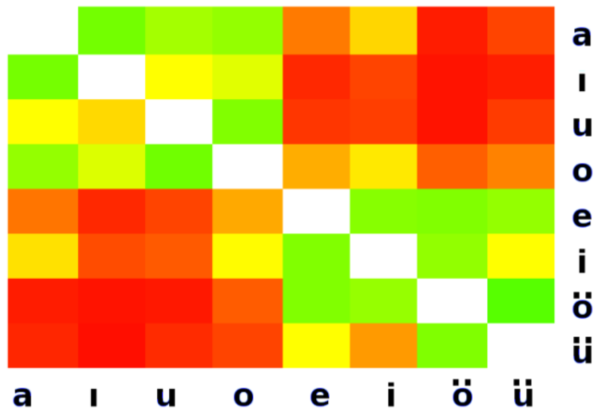

Figure 1.1. Heatmap of likelihood of vowel sequences in Turkish (reproduced from Caplan & Kodner, 2018).

In Figure 1.1, the top right and the bottom left quadrants that are populated by front-back or back-front vowel sequences are mostly darker shades of red, suggesting low likelihood of vowel cooccurrence. In contrast, the top left and bottom right quadrants that are populated by vowel sequences that agree in backness are mostly shades of green, suggesting high likelihood of vowel cooccurrence. However, the shades of the greens vary across the vowel sequences as a function of vowel height and rounding. As such, the distribution of red and green cells in Figure 1.1 closely parallels the distribution of the plus and minus signs denoting vowel sequence permissibility in Table 1.4. The permissibility table and likelihood heatmap diverge in that whereas Table 1.4 reduces vowel cooccurrence to discrete signs, Figure 1.1 represents the granular variability in vowel cooccurrence relationships.

Caplan and Kodner (2018) used k-means clustering to test their model outputs. The Turkish model's k-means clusters revealed that the model predicts front-back and rounded-unrounded harmony classes as well as that nonhigh vowels do not participate in rounding harmony as targets. In contrast, in the language models where there is no vowel harmony such as

English, the model revealed rather uniform likelihoods of vowel cooccurrence that suggest independent distribution of vowels.

Similarly, Baker (2009) analyzes a 100,000-sentence Turkish corpus from the Leipzig University Wortschatz collection to model vowel transitions in a two-state Hidden Markov Model (HMM). The model yielded /ɑ ɯ o u/ and /e i œ y/ as the two groups of vowels: i.e., back and front vowel classes, respectively. Transition probabilities were high within each group and low across the groups. These results suggest that in Turkish, back vowels transition into back vowels and front vowels transition into front vowels. Similar to the Turkish vowel groups, a two-state HMM modeling Italian yielded /a o u/ and /e i/ as the two vowel groups, i.e., back and front vowel classes, respectively. However, the transition probabilities were not as strong and suggested a slightly above chance tendency of alternation between the groups. That is, in Italian, transitions between successive vowels are arbitrary with respect to the two vowel classes. Thus, the HMMs suggest that distributional regularities in corpora can be identified in a way that captures variability and harmony can be deduced from these regularities.

Together, the studies reviewed above (Baker, 2009; Caplan & Kodner, 2018) support the claim that harmony classes might emerge from distributional regularities in the lexicon (Cole, 2009). In addition, Kabak and Weber (2013) present comparisons of expected and observed vowel sequence frequencies, which provide evidence in support of the vowel cooccurrence patterns discussed above, and moreover show that there is an interaction of vowel height and backness. In high vowel sequences, agreement in backness is more frequent than expected and disagreement in backness is much less frequent than expected. This additional pattern suggests that there might be statistical regularities in Turkish vowel cooccurrence that have not been discussed in the literature before.

1.3. Vowel cooccurrence in the Turkish lexicon as reflected in vowel perception

As reviewed above, Turkish vowel harmony incompletely applies to the Turkish lexicon. As a result, some vowel cooccurrence patterns where the vowels do not agree in a number of features are observed. Sometimes, these patterns are more frequent in the lexicon than patterns that are fully harmonic. A number of experimental studies have looked at whether Turkish speakers' expectations and/or preferences for vowel sequences in nonwords are predicted by the exact vowel cooccurrence patterns in the Turkish lexicon or a feature-based, formal description of these patterns. In one such pioneering study, Zimmer (1969) tests native Turkish speakers' awareness of three vowel cooccurrence tendencies observed in the Turkish lexicon. First, vowels cooccurring within a word agree in backness. Second, vowels cooccurring within a word agree in roundness when the second vowel is high. The third tendency is an exception to the second tendency. When the first vowel is /ɑ/ and is followed by a labial consonant, the second vowel is rounded /u/ rather than /ɯ/. This tendency is observed by Lees (1966) and is termed *labial attraction* on the grounds that the second vowel purportedly agrees with the intervening labial consonant rather than the first vowel in rounding. Zimmer asked native Turkish listeners to indicate their preferences between disyllabic nonword pairs that follow these tendencies to varying degrees. Some nonword pairs were designed to test the first two tendencies only and some others to test the 'labial attraction' tendency. In the latter pairs, the first vowel was /ɑ/ and the second vowel was either /ɯ/ or /u/. In some of these pairs, the consonant intervening the two vowels was a labial and in others it was a coronal or a dorsal.

For the nonword pairs testing the first two tendencies only, participants predominantly preferred the nonword in which the vowels agree in backness and roundness (84.0%). This finding suggests that native Turkish listeners prefer nonwords that reflect the vowel cooccurrence tendencies regarding agreement in backness and rounding in the Turkish lexicon. In contrast, participants preferred the nonword in which the vowels agree in backness and roundness to a considerably lesser degree between the nonwords with /ɑ-u/ and /ɑ-ɯ/ vowel sequences (/ɑ-ɯ/: 55.8%). Among these nonword pairs, when the intervening consonant was labial, participants had a preference at chance level (/ɑ-ɯ/: 51.9%). In a footnote, Zimmer reports results from a follow up experiment. According to Zimmer, there was a “strong” preference for /ɯ-ɯ/ vowel sequences over /ɯ-u/ in this experiment in contrast to the former participants not having a preference between /ɑ-u/ and /ɑ-ɯ/ vowel sequences.

Zimmer (1969) mainly interprets the findings from the labial attraction pairs with reference to individual differences. He observes that one labial attraction pair exhibited a unanimous preference for the nonword with the /ɑ-u/ sequence. Zimmer notes that this nonword (/tɑvuz/) is highly similar to an existing Turkish word (/tavus/ ‘peacock’). Based on this observation, Zimmer speculates that participant responses may have been driven by similarity to existing words rather than awareness of distributional tendencies over the entire lexicon. Zimmer argues that participant responses indicate a preference in line with either the agreement in roundness tendency or the exception to this tendency conditioned by the labial consonant. That is, on the whole, native Turkish listeners have awareness of both of these conflicting tendencies. However, participant preferences do not exhibit strict adherence to the consonant-conditioned vowel cooccurrence patterns in the lexicon. Rather, native Turkish listeners might hold more generalized representations of the vowel cooccurrence patterns.

Yavaş (1980) replicates Zimmer (1969) with a focus on agreement in backness. Yavaş presented native Turkish speakers with nonword pairs that differ only in whether the second vowel agrees with the first vowel in backness. The participants were asked to indicate which nonword they thought could be a Turkish word. Yavaş observes that participant preference for agreement in backness across the 12 nonword pairs ranges between clear preference (/CœCeC/ >

/CœCɑC/: 100%, /CɑCuC/ > /CɑCyC/: 96%, and /CeCiC/ > /CeCɯC/: 96%) and no preference (/CeCeC/ > /CeCɑC/: 50%, /CiCeC/ > /CiCɑC/: 50%). That is, formal agreement in vowel features does not predict a clear preference over disagreement. Yavaş argues that this variation in preference for agreement or disagreement in backness is better explained by whether a specific vowel sequence is frequent in the Turkish lexicon.

Zimmer and Küntay (2003) asked native Turkish speakers to indicate which vowel they would pick to fill in /CVC_C/ nonwords to make a plausible Turkish word. The authors found that 80-90% of responses were "harmonic," although they do not explain how they define a "harmonic" response among the 8 Turkish vowels for each of the nonwords. The authors also asked participants to indicate what strategies they used in the experiment. The majority of the participants mentioned that Turkish vowel harmony guided their responses, whereas a smaller number of participants mentioned similarity to existing Turkish words influenced their responses. Although participants referred to formal rules or familiarity with the vowel sequences in the lexicon, the responses of the two groups of participants did not necessarily differ, suggesting that a similar decision mechanism underlies all participants' responses.

Oded, Idsardi, and Rhone (2008) present a 2-alternative forced choice vowel identification study. The stimuli were disyllabic nonwords where the first vowel is /ɑ/ and the second vowel is either /i/ or /u/. Participants were asked to identify the second vowel in the nonwords as /i/ or /u/. The authors analyze incorrect responses as an indicator of listener expectation with two hypotheses that make opposing predictions: a 'vowel harmony' hypothesis and a 'frequency' hypothesis. The 'vowel harmony' hypothesis is based on agreement in vowel features. The vowel /u/ agrees with /ɑ/ in backness whereas /i/ does not. Hence, this hypothesis predicts more /u/ responses than /i/ responses. The 'frequency' hypothesis is based on how frequent vowel sequences are in the Turkish lexicon. The vowel sequence /ɑ-i/ has higher type frequency in the

Turkish lexicon than /ɑ-u/. Hence, this hypothesis predicts more /i/ responses than /u/ responses. The results were in line with the 'frequency' hypothesis. Native Turkish listeners made more incorrect responses whereby they mistakenly identified the second vowel as /i/ more than as /u/. In other words, familiarity with the vowel sequence in the lexicon was more influential than feature-based agreement on vowel perception.

Oded et al. (2008) also comparatively tested native English listeners with the same stimuli. In contrast to Turkish where vowel cooccurrence is governed by feature agreement, English vowel cooccurrence is unconstrained with respect to vowel features. Similar to native Turkish listeners, native English listeners also made more incorrect /i/ responses than incorrect /u/ responses. That is, native English listeners had an expectation for /ɑ-i/ vowel sequences over /ɑ-u/ vowel sequences. This result is similarly predicted by the higher frequency of /ɑ-i/ vowel sequences than /ɑ-u/ vowel sequences in the English lexicon. These findings suggest that listeners predict upcoming segments based on lexical distribution of individual segment sequences regardless of whether more general, feature-based cooccurrence tendencies exist in the lexicon of their native language or not.

In this section, I reviewed several studies examining whether Turkish vowel perception in multisyllabic contexts exhibit a preference for vowels to agree in certain features. Together, these studies show that expectations and preferences in Turkish vowel perception are not solely predicted by agreement in vowel features. The results reviewed suggest that similarity to existing words and frequency of specific vowel sequences in the lexicon also contribute to Turkish vowel perception beyond vowel feature-driven expectations and preferences. That is, these studies show that vowel cooccurrence in Turkish influences Turkish vowel perception. Moreover, this influence is potentially more detailed and granular than feature-level generalizations. Each study reviewed above suggests that existing words or patterns observed in the lexicon might be more influential than agreement in features on speaker-listener expectations and preferences for vowel cooccurrence in Turkish.

1.4. Quantifying vowel cooccurrence patterns in the lexicon

As suggested by the findings reviewed above, vowel cooccurrence patterns in the lexicon is part of speakers-listeners' phonological and phonotactic knowledge in Turkish. In the present study, I aim to explore whether lexical vowel cooccurrence influences perceived vowel similarity whereby higher likelihood of vowel cooccurrence leads to higher perceived similarity. Following previous studies that computationally model harmony classes as emergent representations from vowel cooccurrence patterns in the Turkish lexicon (e.g., Baker, 2009), I use Pointwise Mutual Information (PMI) to quantify likelihood of vowel cooccurrence in the lexicon and as a measure of phonotactic similarity of vowels. I calculate PMI as in (3) , where *a* and *b* are two vowels, $P(a)$ and $P(b)$ are the unigram probabilities of the two vowels in disyllabic words, and $P(a|b)$ and $P(b|a)$ are the bigram probabilities of the vowel sequences /bC$^+$a/ and /aC$^+$b/, respectively.

(3) $PMI(a,b) = \log_2 \frac{P(b|a)+P(b|a)}{P(a)\times P(b)}$

The multiple of the unigram probabilities of the two vowels yields the expected probability of *a* and *b* cooccurring. To compare the expected and observed probabilities, the bigram probability of *a* and *b* cooccurring is divided by the expected cooccurrence probability. This resulting number's logarithm to the base of 2 yields a comparable measure of observed as opposed to expected probabilities across varying vowels. A PMI value of 0 indicates that the vowel pair cooccurs at the expected rate based on unigram probabilities of each of the vowels, suggesting that the two vowels are independently distributed in the lexicon with no vowel cooccurrence restrictions. Positive PMI values indicate that the vowel pair is more likely to cooccur than would be expected based on their unigram probabilities. Conversely, negative PMI values indicate that the vowel pair is less likely to cooccur than would be expected based on their unigram probabilities. That is, harmonic vowels are predicted to have positive PMI values whereas disharmonic vowels are predicted to have negative PMI values. A greater absolute PMI value indicates a stronger likelihood of cooccurrence relationship.

Figure 1.2 visualizes in a heatmap the PMI values that I calculated over 18,166 disyllabic Turkish words from the 1998 Turkish Dictionary in the Turkish WordNet corpus (Bakay et al., 2021). In this heatmap, purple cells represent vowels with positive PMI values and high likelihood of cooccurrence and red cells represent vowels with negative PMI values and low likelihood of cooccurrence. Darker shades represent a stronger divergence from expected probability under independent distribution of vowels. Darker purple indicates a stronger tendency for the vowels to cooccur whereas darker red indicates a stronger tendency for the vowels to not cooccur.

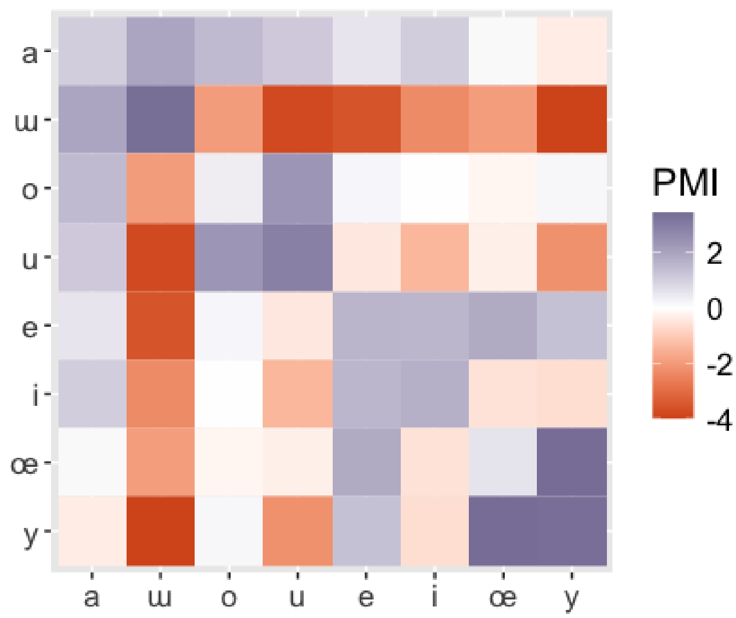

Figure 1.2. Heatmap of likelihood of vowel cooccurrence in the Turkish lexicon as measured by PMI.

Notice that the top left and the bottom right quadrants in Figure 1.2 have relatively more purple cells than the top right and bottom left quadrants. This pattern is in line with backness harmony in Turkish, as the vowel pairs in the top left and bottom right quadrants agree in backness. Also notice the red cells within the top left and the bottom right quadrants where vowels agree in backness and yet disagree in rounding with a high vowel. These observations show that PMI calculated over the Turkish lexicon yields a description of vowel cooccurrence that is aligned with previous discrete descriptions of Turkish vowel harmony. What PMI captures in addition is that harmony applies incompletely in the Turkish lexicon and that both harmonic and disharmonic vowel pairs vary in their likelihood of cooccurrence.

Consider, for instance, the vowels /ɑ/ and /ɯ/. These vowels agree in backness and rounding and hence are harmonic. The vowel pairs /ɑ/-/ɑ/, /ɑ/-/ɯ/, and /ɯ/-/ɯ/ all have positive PMI values as reflected in the cells being purple, yet they vary in the exact PMI values and in their likelihood of cooccurrence in the lexicon. The vowel pair with the highest likelihood of cooccurrence among these three pairs is the /ɯ/-/ɯ/ pair as indicated by the darker shade of purple relative to /ɑ/-/ɑ/ and /ɑ/-/ɯ/ pairs. In other words, /ɯ/-/ɯ/ is more likely to cooccur in the Turkish lexicon compared to /ɑ/-/ɑ/ and /ɑ/-/ɯ/, despite all pairs being equally harmonic in terms of discrete vowel feature agreement.

The predictions that follow from the above observations about the vowel cooccurrence patterns in the Turkish lexicon are twofold. First, vowels that are likely to cooccur in the lexicon are predicted to be coactivated more and hence be perceived to be more similar than vowels that are likely to not cooccur. This prediction does not necessarily rely on a gradient description of vowel cooccurrence patterns in the lexicon. Rather, a discrete description of Turkish vowel harmony would also lead to the same prediction under the assumption that vowels that are harmonic and hence cooccur are coactivated. I argue with others that this pattern of coactivation would lead to emergent harmony classes (e.g., Baker, 2009; Caplan & Kodner, 2018; Cole, 2009). The second prediction follows from the observations that vowel harmony incompletely applies and there is variability in vowel cooccurrence patterns in the lexicon. This variability is captured by a measure such as PMI but not by a discrete description of vowel harmony. Given this variability, vowels that cooccur with greater likelihood in the lexicon would also be more likely to be coactivated. By extension, higher likelihood of vowel cooccurrence would lead to higher perceived vowel similarity relative to vowels that cooccur with lesser likelihood. That is, within a harmony class, vowels that are more likely to cooccur relative to other vowels would also be perceived to be more similar. As a result, I hypothesize that PMI as a continuous measure of likelihood of vowel cooccurrence will be a predictor of perceived vowel similarity in Turkish beyond phonetic similarity of vowels. I test this hypothesis in a series of vowel discrimination and identification experiments, as detailed in the next section.

1.5. The present experiments

The goal of this dissertation is to examine the predictors of perceived vowel similarity in Turkish and in particular likelihood of vowel cooccurrence in the Turkish lexicon as a predictor. Following previous studies (e.g., Hall & Hume, 2015; see Chapter 4), I predict that phonetic similarity of stimuli will be the strongest predictor of perceived vowel similarity overall. To better understand phonetic similarity of vowels in Turkish and explore the Turkish vowel space in general, I first present a Turkish vowel production study in Chapter 2. My main research questions in this vowel production study are how Turkish vowel similarity can be described based on vowel phonetics and how phonetic and phonological variation in Turkish vowel production relates to vowel features and/or harmony classes. To this end, I analyze the Turkish vowel productions with respect to phonetic distance and overlap in the vowel space. I also examine whether Turkish vowel features defined phonologically are phonetically grounded and whether vowels sharing features behave similarly in other phonological contexts.

In Chapters 3 and 4, I present Turkish vowel perception studies exploring vowel similarity as perceived by native speakers. More specifically, Chapter 3 presents two Turkish vowel discrimination experiments and two English vowel discrimination experiments for a cross-linguistic comparison. My research questions in the vowel discrimination experiments are how various phonetic and phonological measures of vowel similarity predict perceived vowel similarity, and in particular what the relative contributions are of these predictors. I analyze response times as a measure of perceived vowel similarity with slower responses indicating higher perceived vowel similarity. I manipulate the task demands in the two experiments to primarily target either the acoustic-phonetic or the phonological processing of the stimuli (e.g., Davidson & Shaw, 2012). I predict greater influence of phonological measures of vowel similarity on response times when the task demands encourage phonological processing of the stimuli. I also present a cross-linguistic comparison with English vowel discrimination, where I predict greater influence of vowel features and the likelihood of vowel cooccurrence on perceived vowel similarity in Turkish than in English.

Chapter 4 presents a Turkish vowel identification experiment. Phonemic vowel identification is assumed to reflect phonological processing to a greater degree relative to tasks where phonemic identification is not strictly necessary, such as vowel discrimination. I analyze perceptual confusion rates between vowels in the vowel identification experiment, with higher rates of confusion indicating higher perceived vowel similarity. I hypothesize that perceptual confusion rates will be predicted by phonetic and phonological measures of vowel similarity as well as likelihood of vowel cooccurrence as a measure of phonotactic similarity of vowels. More specifically, I hypothesize that higher likelihood of vowel cooccurrence will predict higher rates of perceptual confusion between vowels.

To preview the results, I find in the Turkish vowel production experiment (Chapter 2) that vowels within a harmony class do not necessarily exhibit similar patterns of phonetic variation or uniformly participate in phonological processes. Rather, I find evidence suggesting that phonetic variation and participation in phonological processes might be conditioned by whether vowels are situated towards the periphery of the vowel space with relatively higher phonetic distinction due to relatively more extreme F1-F2 values (see below). In the Turkish vowel discrimination experiments (Chapter 3), I find that listener performance varies across tasks that prioritize either acoustic-phonetic or phonological processing of the stimuli. In the former, perceived vowel similarity is predicted by the degree of overlap between vowel categories in the vowel space. In the latter, perceived vowel similarity is predicted by dissimilarity in vowel features. These results suggest that phonological vowel processing influences perceived vowel similarity beyond the phonetic similarity of vowels. That is, similarity-based coactivation in the exemplar space involves phonological knowledge. In the Turkish vowel identification experiment (Chapter 4), I find that perceived vowel similarity is predicted by a combination of phonetic and phonological measures of vowel similarity including likelihood of vowel cooccurrence in the Turkish lexicon. However, the effect of phonotactic similarity is not in the expected direction. Based on some observations in the literature on Turkish vowel harmony being conditioned by vowel peripherality and perceptual confusability of vowels (e.g., Clements & Sezer, 1982; Suomi, 1983), I present a post-hoc analysis that adds vowel peripherality as a predictor. To anticipate the discussion, I review these accounts that view Turkish vowel harmony from the perspective of vowel peripherality in the following section.

1.6. Vowel peripherality and Turkish vowel harmony

As described above, the vowel production and perception experiments were designed to explore how a number of phonetic, phonological, and phonotactic factors contribute to perceived vowel similarity in Turkish. The results of the vowel production (Chapter 2) and vowel identification (Chapter 4) experiments led me to consider vowel peripherality in Turkish as it relates to Turkish vowel harmony and Turkish vowel perception. In the vowel identification experiment in Chapter 4, I ran post-hoc analyses with vowel peripherality as an additional predictor. To anticipate the discussion and the post-hoc analyses, I first provide an overview of how vowel peripherality might influence perceived vowel similarity and how Turkish vowel peripherality has been treated in the literature on Turkish vowel harmony.

Vowel peripherality is a property of vowels in terms of where they are situated in the phonetic vowel space. Simply put, vowels that are towards the outer edge of the F1-F2 space are peripheral as opposed to nonperipheral (e.g., Labov, Yaeger, & Steiner, 1972; Polka & Bohn, 2003). Similarly, peripheral vowels are also defined by converging or focalized F1-F2 values (e.g., Masapollo, Polka, Molnar, & Ménard, 2017; Schwartz, Abry, Boë, Ménard, & Vallée, 2005). Labov (1994) argues that among vowels on the same height level, vowels that are less peripheral are less distinct and nonoptimal. Of interest here is whether vowel peripherality influences vowel perception and in particular perceived vowel similarity considering phonetic distinctiveness is one of its correlates.

In vowel identification, previous research suggests that peripheral vowels are identified with higher accuracy whereas nonperipheral vowels exhibit more perceptual confusions (e.g., Hall & Hume, 2015; Hillenbrand, Getty, Clark, & Wheeler, 1995; Peterson & Barney, 1952). In infant vowel discrimination, vowel peripherality is found to yield perceptual asymmetries suggesting that peripheral vowels have a perceptual advantage over nonperipheral vowels (e.g., Masapollo et al., 2017; Polka & Bohn, 2003; 2011). Findings suggest that discrimination is easier when the nonperipheral vowel is presented first than when the peripheral vowel is presented first. In other words, a peripheral vowel is perceived to be more dissimilar from a nonperipheral vowel than vice versa. This perceptual asymmetry in vowel discrimination is attributed to peripheral vowels being ‘natural referents’ or perceptual anchors with converging vowel formants that make them perceptually more prominent (Polka & Bohn, 2003; 2011). However, in adult vowel perception, the perceptual asymmetry is limited to nonnative vowels. No perceptual asymmetry is observed in discrimination between peripheral and nonperipheral vowels that are phonemic in the listeners’ native language (Polka & Bohn, 2003). On the other hand, a perceptual asymmetry is observed in native vowel discrimination within phonemes when the less peripheral token is presented first (Masapollo et al., 2017).

Paralleling the ideas about phonetic and perceptual distinctiveness as the hallmark of vowel peripherality, Suomi (1983) proposes a similar classification. Based on cross-linguistic observation, Suomi claims that the vowel set /ɑ e i o u/ is defined by relatively extreme F1-F2 values and are more common across vowel inventories due to being perceptually maximally distinct. Suomi terms these vowels as ‘strong’ whereas other vowels with less extreme F1-F2 values that are perceptually less distinct are termed ‘weak.’ As such, Suomi’s classification of vowels as perceptually strong and weak is aligned with the phonetic classification of vowels as peripheral and nonperipheral. Similar to the work reviewed above, Suomi argues that weak (nonperipheral) vowels are perceptually compromised due to their relative lack of phonetic definition. Suomi’s original contribution is the proposal that Turkish vowel harmony is perceptually motivated to help the accurate identification of the perceptually weak nonperipheral Turkish vowels. More specifically, vowel harmony restricts vowel cooccurrence for noninitial vowels, thereby increasing the predictability of noninitial vowels within a word based on the initial vowel. According to Suomi, the motivation to enhance the accurate identification of nonperipheral vowels in noninitial positions have prompted the vowel cooccurrence restrictions in Turkish.

Suomi (1983) argues that weak vowels are especially prone to perceptual confusions with the neighboring strong vowels that are on the same F1 level as the weak target vowel but have more extreme F2 values. In other words, weak vowels are perceptually compromised due to being insufficiently distinct on the F2 plane. In contrast, strong vowels are more prominent to the disadvantage of the neighboring weak vowels on the same F1 level. In Turkish, vowels make three F1 levels (see Kopkallı-Yavuz, 2010 and Chapter 2): /i y ɯ u/ are high, /e œ o/ are mid, and /ɑ/ is low. Accordingly, /ɯ œ y/ are the weak vowels of Turkish whereas /ɑ e i o u/ are strong. Based on the perceptual confusability claim stated above, the weak vowels /ɯ/, /œ/, and /y/ are predicted to be perceptually highly confusable with the strong vowels /u/, /e/, and /i/, respectively. According to Suomi, Turkish vowel harmony is motivated to reduce these perceptual confusions that are due to contrasts along the F2 dimension not being optimal.

Suomi's argument is that Turkish vowels in noninitial position can be accurately identified by its F1 level alone by virtue of Turkish vowel harmony. That is, Turkish vowel harmony increases the predictability of a noninitial vowel's identity without having to estimate its F2. As vowel F2 is perceptually less distinct for nonperipheral vowels, it is nonperipheral vowels that benefit from increased predictability.

Turkish vowel harmony increases the predictability of noninitial vowels as follows. Suomi describes Turkish vowels in four F2 levels as a combination of backness and rounding: /e i/ are high F2, /œ y/ are high-mid F2, /ɯ ɑ/ are low-mid F2, and /u o/ are low F2. The only F1 level that has vowels of all F2 levels is the high F1 level. Two of the three nonperipheral Turkish vowels are also on the high F1 level. Fittingly, the vowels on the high F1 level are subject to rounding harmony in noninitial positions in addition to backness harmony. As such, in noninitial positions, the F2 level of a high-F1 vowel can be fully predicted by the initial vowel's backness and rounding. On the other hand, the F2 level of a nonhigh-F1 vowel can be fully predicted by the initial vowel's backness alone as nonhigh rounded vowels do not occur in noninitial positions. As such, vowel cooccurrence restrictions in Turkish make the F2 levels of noninitial vowels of any F1 level contextually predictable. Suomi proposes that reducing the cue weight of F2 of the noninitial vowel is for the benefit of nonperipheral vowel identification, as peripheral vowels have prominent F2 values unlike nonperipheral vowels.

Following Suomi (1983) and based on observations from rounding harmony, Kaun (2004) similarly proposes that harmony is a perceptually motivated phenomenon. Kaun argues that harmony extends a "perceptually vulnerable quality" or feature across segments. The listener is provided with cues to a feature in the trigger as well as in the target vowel. Thus, the probability that the feature will be accurately identified is increased. Converging with Suomi, Kaun argues that harmony is motivated to perceptually enhance contrasts that are prone to misidentification.

Koo and Cole (2007) present a series of experiments testing a similar idea. In their view, a target sound is identified probabilistically based on its phonetic configuration and, secondarily, based on the context in which the sound occurs. The context assigns identification probabilities to viable alternatives based on the frequency of each sound pattern in the lexicon. As such, phonotactics are reflected in probabilistic speech perception. Koo and Cole's "perceptual facilitation hypothesis" indicates that phonotactic knowledge will have a greater influence on perception when the target sound is perceptually more confusable. They test this hypothesis in artificial grammar learning experiments.

In their experiment, participants were exposed to (nonsense) words with a phonotactic constraint. For example, participants were trained on a phonotactic constrained that allowed words such as /sa.la.la/ and /sa.ra.ra/ but not /sa.la.ra/ or /sa.ra.la/. After exposure, participants were asked to repeat novel nonwords that either followed the same phonotactic constraint (e.g., /ke.la.la/) or did not (e.g., /ke.la.ra/). Participants were exposed to the phonotactic constraint with perceptually more confusable phonemes (/l-r/) or perceptually less confusable phonemes (/l-m/). Participants who learned the phonotactic constraint with the perceptually more confusable phonemes were faster in accurately repeating the words that conform to the phonotactic constraint compared to words that violate the phonotactic constraint. In contrast, no difference in response times was observed for participants who learned the phonotactic constraint with perceptually less confusable phonemes. That is, learning a phonotactic constraint facilitated perception and/or production when the phonotactic constraint involved perceptually more confusable phonemes. In contrast, there was no facilitation when the phonotactic constraint learned involved perceptually less confusable phonemes. Koo and Cole (2007) also present a connectionist model that predicts their experimental results. They conclude that phonotactic knowledge can reduce perceptual confusions.

Suomi (1983)'s claim operates under a rule-based and rather exceptionless description of Turkish vowel harmony. As such, according to Suomi, Turkish vowel harmony applies to perceptually weak and strong vowels alike, yet is of greater benefit to the perceptually weak Turkish vowels /ɯ œ y/. However, not all Turkish words are harmonic. Suomi's claim is that harmonic contexts benefit perceptually weak vowels. By extension, in disharmonic contexts, perceptually weak vowels would be perceptually disadvantaged to a greater degree than perceptually strong vowels. Fittingly, Clements and Sezer (1982) observe that in Turkish, /ɑ e i o u/ are attested in disharmonic Turkish words more frequently than /ɯ œ y/. In other words, Clements and Sezer claim that there is a greater degree of vowel cooccurrence restrictions on perceptually weak, nonperipheral Turkish vowels such that they occur in contextually predictable environments. On the other hand, perceptually strong, peripheral Turkish vowels do not rely on contextual predictability for accurate identification and hence are allowed to occur in disharmonic contexts to a greater degree.

Together, the works of Suomi (1983) and Clements and Sezer (1982) reviewed above remark that nonperipheral Turkish vowels are perceptually weak and yet perceptually enhanced in harmonic contexts, and that nonperipheral vowels are confined to predictable contexts in the Turkish lexicon. Several Turkish vowel perception studies present results that can be interpreted in relation to these claims.

That /ɯ œ y/ are perceptually weak finds support in a synthesized Turkish vowel identification study: Yılmaz (2022) finds that nonperipheral Turkish vowels are numerically less accurately identified (/ɯ/: 59.3%, /œ/: 54.9%, /y/: 53.4%) than peripheral Turkish vowels (/ɑ/: 99%, /e/: 85.8%, /i/: 98.5%, /o/: 79.4%, /u/: 87.7%). I return to this observation in Chapter 4 in my analysis of the naturally produced Turkish vowel identification data. In addition, as reviewed above, Yavaş (1980) found that native Turkish listeners favored disharmonic nonwords with vowels were from the /ɑ e i o u/ vowel set in acceptability ratings. Clements and Sezer (1982) interpret this finding to suggest that disharmonic vowel pairs consisting of the /ɑ e i o u/ vowel set are in fact well-formed. The higher acceptability ratings might also reflect that these disharmonic vowel pairs are attested more frequently in the Turkish lexicon, as claimed by Clements and Sezer.

To my knowledge, Kabak and Weber (2013) is the only corpus study that tests Clements and Sezer's (1982) claim that peripheral Turkish vowels are more frequently attested in disharmonic words. Their analysis compares the frequencies of disharmonic vowel sequences in a corpus of Turkish words. Kabak and Weber show that disharmonic sequences where both vowels are from the vowel set /ɑ e i o u/ are generally more prevalent than disharmonic sequences one of the vowels is from the vowel set /ɯ œ y/. However, they also note some individual vowel sequence frequencies that go against Clements and Sezer's generalizations. I replicate Kabak and Weber's analysis in a linear regression. I model PMI as a measure of likelihood of vowel cooccurrence for disharmonic Turkish vowel pairs with vowel peripherality as the predictor. The results support the observation that in disharmonic contexts, peripheral vowels are significantly more likely to cooccur with other peripheral vowels (estimated mean PMI = 0.41, $SE = 0.25$) than with nonperipheral vowels (estimated mean PMI = -1.08, $SE = 0.31$; $\beta = 0.74$, $SE = 0.20$, $t = 3.71$, $p = 0.002$). In other words, I find evidence supporting that nonperipheral vowels are subject to a greater degree of cooccurrence restrictions than peripheral vowels.

The observations reviewed above suggest that vowel peripherality might be a predictor of vowel identification accuracy in Turkish such that peripheral vowels are identified with higher accuracy (Hall & Hume, 2015; Hillenbrand et al., 1995; Suomi, 1983; Yılmaz, 2022). Moreover, Turkish vowel harmony might be perceptually motivated to increase the identification accuracy of nonperipheral Turkish vowels by restricting these vowels to more predictable contexts

(Clements & Sezer, 1982; Kabak & Weber, 2013; Kaun, 2004; Koo & Cole, 2007; Suomi, 1983). As such, considerations of vowel peripherality presents an alternative view of Turkish vowel harmony. Turkish vowel harmony might serve a function to reduce perceptual confusions for nonperipheral vowels. I return to this proposal in Chapter 4 for an alternative interpretation of the perceptual confusion results in Turkish vowel identification previewed above.

CHAPTER 2

TURKISH VOWEL PRODUCTION

In order to gain a better understanding of the Turkish vowel space and how it relates to Turkish vowel harmony, I ran a Turkish vowel production study in which I collected vowel productions in controlled contexts from native Turkish speakers. To my knowledge, the largest Turkish vowel production data to date includes vowel productions from 52 native Turkish speakers, 23 of whom are 6-year-old children (Türk et al., 2004). However, the majority of the vowels in that study were produced in sentences in which phonetic context was not controlled, which may have introduced undue variability. On the other hand, Turkish vowel production studies in which more control over vowel context is exerted either have smaller numbers of tokens produced by smaller number of speakers or focus on a specific vowel and context. For instance, Kopkalli-Yavuz (2010) conducted a Turkish vowel production study in which vowels were produced in controlled /kVp/ environments, however, there were only 7 speakers who produced 5 repetitions of each stimulus. Moreover, these previous Turkish vowel production studies that describe the vowel space do not provide detailed phonetic analyses that explore the interaction of phonetics and phonology. The present study aims to collect a larger data set of Turkish vowels produced by a higher number of speakers in controlled environments to provide a more detailed analysis of the vowel space, including coarticulatory and allophonic variation, and explore how these phenomena relate to Turkish phonology. More specifically, I ask whether vowels are well separated in the vowel space based on vowel features or harmony classes with greater phonetic distance and higher dissimilarity between disharmonic vowels than between harmonic vowels, and whether vowels within the same harmony class behave similarly under similar phonological conditions such as contexts that license allophonic variation.

2.1. Methods

2.1.1. Participants

Twenty-six native, monolingually raised Turkish speakers aged 19-56 (M_{age} = 28.46, 17 female) participated in the production study. Participants were Northwestern University affiliates (students or researchers) and their family members. All participants reported growing up in Turkey in Turkish speaking families. All participants were self-reported fluent L2 speakers of English. Participants' mean self-reported age of onset for English was 9 years of age (range: 2-18), and no participant reported having lived in an English speaking family. Nineteen participants reported knowing a third language, 11 participants reported knowing a fourth language, and 2 participants reported knowing a fifth language. Participants were born and raised in Turkey and had been living in the US at the time of their participation in the study for 2 months-9 years (M_{years} = 3.52). Participants were paid $15 USD as compensation.

2.1.2. Stimuli

Stimuli consisted of two sets of nonword stimuli (monosyllabic and disyllabic nonwords), two sets of Turkish word stimuli (/kVp/ minimal pair words replicating Kopkalli-Yavuz, 2010 and 168 Turkish words representative of a variety of contextual factors), and two short passages.

Monosyllabic nonword stimuli consisted of the 8 Turkish vowels (/ɑ e ɯ i o œ u y/) in /bVb/ nonword context, embedded in a carrier sentence (*Demet /bVb/ dedi* 'Demet /bVb/ said'). This nonword context was chosen to minimize consonant-to-vowel and vowel-to-vowel coarticulation. Disyllabic nonword stimuli consisted of the 8 Turkish vowels (/ɑ e ɯ i o œ u y/) in V_1 and V_2 positions in /bV_1b.bV_2b/ nonword context, exhausting all V_1-V_2 combinations, embedded in a carrier sentence (*Demet /bV_1bbV_2b/ dedi* "Demet /bV_1bbV_2b/ said"). This nonword context was chosen to examine vowel-to-vowel coarticulation while minimizing consonantal coarticulatory effects.

To complement the monosyllabic nonwords, a set of real Turkish minimal pair words were elicited following Kopkalli-Yavuz (2010). The minimal pair word stimuli consisted of the 8 Turkish vowels (/ɑ e ɯ i o œ u y/) in /kVp/ word context, embedded in a carrier sentence (*Demet /kVp/ dedi* "Demet /kVp/ said"). This word context was chosen to provide a baseline for comparisons across monosyllabic Turkish words. Turkish word stimuli consisted of 106 monosyllabic Turkish words, 60 disyllabic Turkish words, and 2 trisyllabic Turkish words, totaling 232 vowel tokens, embedded in a carrier sentence (*Demet 'target word' dedi* "Demet 'target word' said"). These words had the 8 Turkish vowels in vowel lengthening, vowel fronting, vowel nasalization, and vowel lowering contexts, as will be described in detail below.

Short passages consisted of "Yalancı Çoban (The Boy Who Cried Wolf)" (Aksu, 2023) and "Jale'nin Dünyası" (Cangökçe Yaşar et al., 2019), which are phonetically near-balanced Turkish passages with 115 unique Turkish words containing 294 vowels and 93 unique Turkish words containing 256 vowels, respectively.

2.1.3. Procedure

The production study was conducted in Northwestern University's Linguistics Laboratories on the Evanston campus in Illinois. Upon arriving at the laboratory for their scheduled recording sessions, participants first gave their informed consent and completed a language background survey. Participants completed the recording session individually in a soundproof booth. They were seated in front of a screen that visually presented the stimuli embedded in carrier sentences or the short passages over a PowerPoint presentation with a keyboard and a mouse to advance trials on their own pace. A Shure SM27 mounted microphone with a windscreen was placed in front of the participant for digital recording of the productions. Participants were instructed to produce the sentences with words and nonwords as well as the passages in a natural and clear speaking style. The production instruction was given verbally before the recording session and visually before every block in the recording session. Participants were also instructed that they could take breaks during the recording session.

Participants were first presented with the /kVp/ minimal pair words, starting with 4 practice trials, after which 3 repetitions of the 8 stimuli were presented in randomized order, totaling 24 productions with 3 tokens for each vowel. The /kVp/ minimal pair word block was followed by the first Turkish word block in which the 168 Turkish words were presented once in a semi-randomized order in which words that are minimal or near minimal pairs did not appear in consecutive trials. The first short passage "Yalancı Çoban" followed the first Turkish word block. Next, the monosyllabic nonword /bVb/ block followed, starting with 4 practice trials, after which 10 repetitions of the 8 stimuli were presented in randomized order, totaling 80 productions with 10 tokens for each vowel. Following the monosyllabic nonword block, participants were presented with the first disyllabic nonword block, beginning with 4 practice trials. Disyllabic nonwords were blocked by V_1 vowel. Each block had 3 repetitions of the 8 stimuli presented in randomized order, totaling 24 productions. The first disyllabic nonword block had 4 V_1 vowel blocks. The second short passage "Jale'nin Dünyası" followed the first disyllabic nonword block, and was in turn followed by the second disyllabic nonword block. The second disyllabic nonword block consisted of the remaining 4 V_1 vowel blocks, and was followed by the second Turkish word block in which the 168 Turkish words were presented once in a semi-randomized order in which words that are minimal or near minimal pairs did not appear in consecutive trials.

2.1.4. Data processing

Audio recordings and TextGrids marking trial intervals with associated sentence transcripts were submitted to MFA for segmentation and forced alignment. Although there is a Turkish MFA acoustic model, due to higher training data, the English MFA acoustic model was used by re-coding the Turkish MFA dictionary to be compatible with the English MFA acoustic model. For 5 of the 8 Turkish vowels (/ɑ e i o u/) and 4 vowel allophones (/ɛ ɪ ɨ ʊ/) that were present in the Turkish dictionary, there is a match between Turkish and English MFA acoustic model vowel categories, and no re-coding was necessary. For the remaining 3 Turkish vowels /ɯ œ y/, there is no directly matching vowel in the English MFA acoustic model. Due to the centrality of these vowels, they were encoded as schwa /ə/. However, as the Turkish vowels are not reduced unlike schwa /ə/, to discourage MFA from assigning reduced schwas to full Turkish vowels, alternative pronunciations with two consecutive schwas /ə ə/ were added to the dictionary as a replacement for the vowels /ɯ œ y/. Lastly, for long Turkish vowels, the vowel length diacritic ː was utilized for long /ɑ e i o u/ vowels and two consecutive schwas /ə ə/ were used for long /ɯ œ y/ vowels. The audio recordings and TextGrids with alignments from MFA were submitted to a Praat script that extracted vowel F1-F4 at five timepoints (20%, 35%, 50%, 65%, and 80% of the vowel) and vowel duration from the target words. Formant ceiling was set to 5000 Hz for men and 5500 Hz for women speakers in the process of formant estimation, and was modified to 4500, 5000, 5500, or 6000 Hz if measurement errors were detected following visual inspection of the formant tracks plotted at the five timepoints.

2.2. Canonical Turkish vowels

Turkish has 8 canonical vowel phonemes, although there is disagreement in the literature with respect to the phonetic classification of these 8 vowels (see Kopkalli-Yavuz, 2010 for a review). In their Handbook of the International Phonetic Association chapter on Turkish, Zimmer and Orgun (1992) posit that the 8 canonical Turkish vowel phonemes are /a e i o u ɯ œ y/. The authors provide a vowel quadrilateral (reproduced in Figure 2.1), which suggests that they classify /i y ɯ u/ as high, /e œ o/ as mid, and /a/ as low in terms of vowel height, and classify /i e y œ/ as front, /a/ as central, and /ɯ u o/ as back in terms of vowel backness. This vowel quadrilateral contrasts with the phonological description of the Turkish vowel system in terms of vowel height and backness in the literature (Table 2.1), which reflects how Turkish vowel phonology treats Turkish vowels for the purposes of vowel harmony (Kornfilt, 1997).

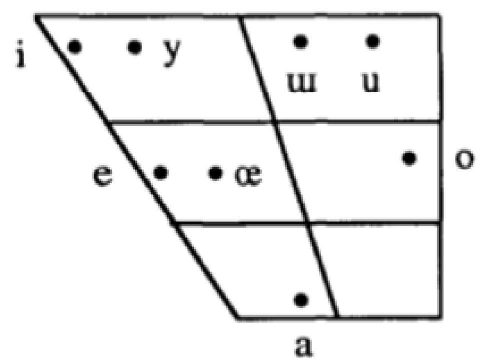

Figure 2.1. Turkish vowel quadrilateral. Reproduced from Zimmer and Orgun (1992).

Table 2.1. Turkish vowel system. Reproduced from Göksel and Kerslake (2005).

	Front		Back	
	Unrounded	Rounded	Unrounded	Rounded
High	i	y	ɯ	u
Non-high (mid and low)	e	œ	a	o

To my knowledge, the single work in the literature that is concerned with phonetically classifying the 8 canonical Turkish vowels with substantial phonetic analysis is the production study by Kopkalli-Yavuz (2010). Kopkalli-Yavuz identifies the 8 canonical vowels of Turkish as /ʌ ɛ i o u ɯ œ y/. According to Kopkalli-Yavuz, phonetically, /i y ɯ u/ are high, /ɛ œ o/ are mid, which are higher than /ʌ/, which is nevertheless not a low but a mid vowel phonetically. Kopkalli-Yavuz also argues that /ʌ o u ɯ/ are phonetically back whereas /ɛ i œ y/ are phonetically front, and that the differences in F2 among the back and front vowels with /o u/ being more back than /ʌ ɯ/ and /ɛ i/ being more front than /œ y/ are accounted for by lip rounding which lowers F2 in vowels. Although I am not necessarily aiming to phonetically classify Turkish vowels, examining how the Turkish vowel phonetics relate to Turkish vowel phonology is one of the motivations of the present study. In the next section, I will present a phonetic analysis of the canonical Turkish vowels in monosyllabic /bVb/ nonword contexts, which control for consonant-to-vowel coarticulation and eliminate vowel-to-vowel coarticulation. I ask whether the canonical Turkish vowels are unambiguously separable to two height classes (high and non-high) and two backness classes (front and back) in the vowel space.

2.2.1. Turkish vowels in monosyllabic nonwords

I analyzed a total of 2077 vowel tokens that were produced in monosyllabic /bVb/ nonwords by the 26 native Turkish speakers to examine the 8 canonical Turkish vowels (Figure 2.2). The phonetic measures of interest were vowel duration (log-transformed) and vowel F1 and F2 at vowel midpoint, which were modeled as the dependent variable using linear mixed effects models with vowel category and speaker gender as predictors. There were random intercepts for speaker but not item due to convergence issues (4). Multi-level fixed effects were evaluated via a type III ANOVA using Satterthwaite's Method. Vowel category was forward difference coded with a specific contrast scheme in each model to reflect the ordering of the vowels for that phonetic measure of interest. Planned comparisons between levels of a variable were performed using the `emmeans` package (Lenth & Piaskowski, 2025).

(4) vowel duration or formant ~ vowel category
+ speaker gender
+ (1 | speaker)

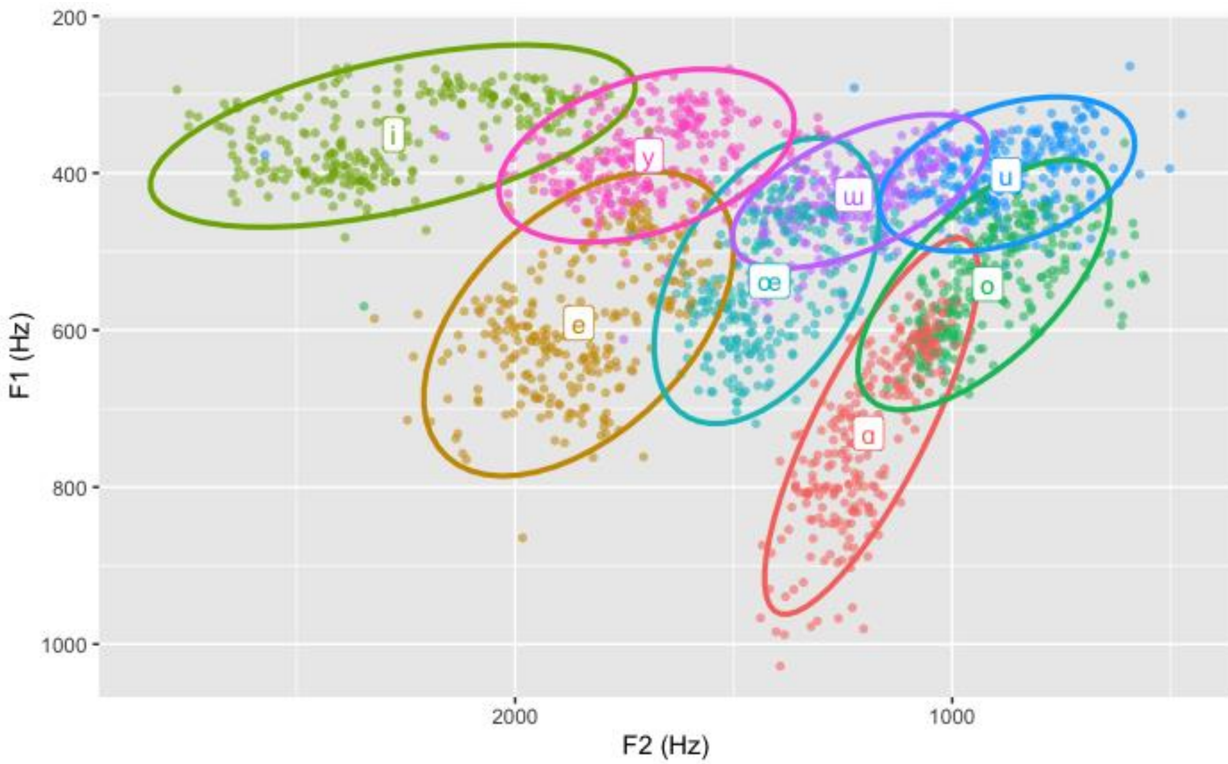

Figure 2.2. Vowel ellipses depicting the vowels produced in monosyllabic nonwords. Each dot represents a single vowel token's raw F1 and F2 measured at vowel midpoint.

In the duration model, there was a significant main effect of vowel category on vowel duration ($F(7, 2051) = 485.02$, $p < 0.0001$), indicating that vowel duration varied across the vowel categories. There were significant differences in vowel duration between all vowel pairs (all *p*s < 0.004; Figure 2.3) except for between /œ/ and /ɑ/ ($p = 0.88$) and between /ɯ/ and /i/ (*p* =

0.71; see Table 2.2). These results are comparable to previous work showing that high vowels are shorter than non-high vowels in Turkish (Şayli, 2002).

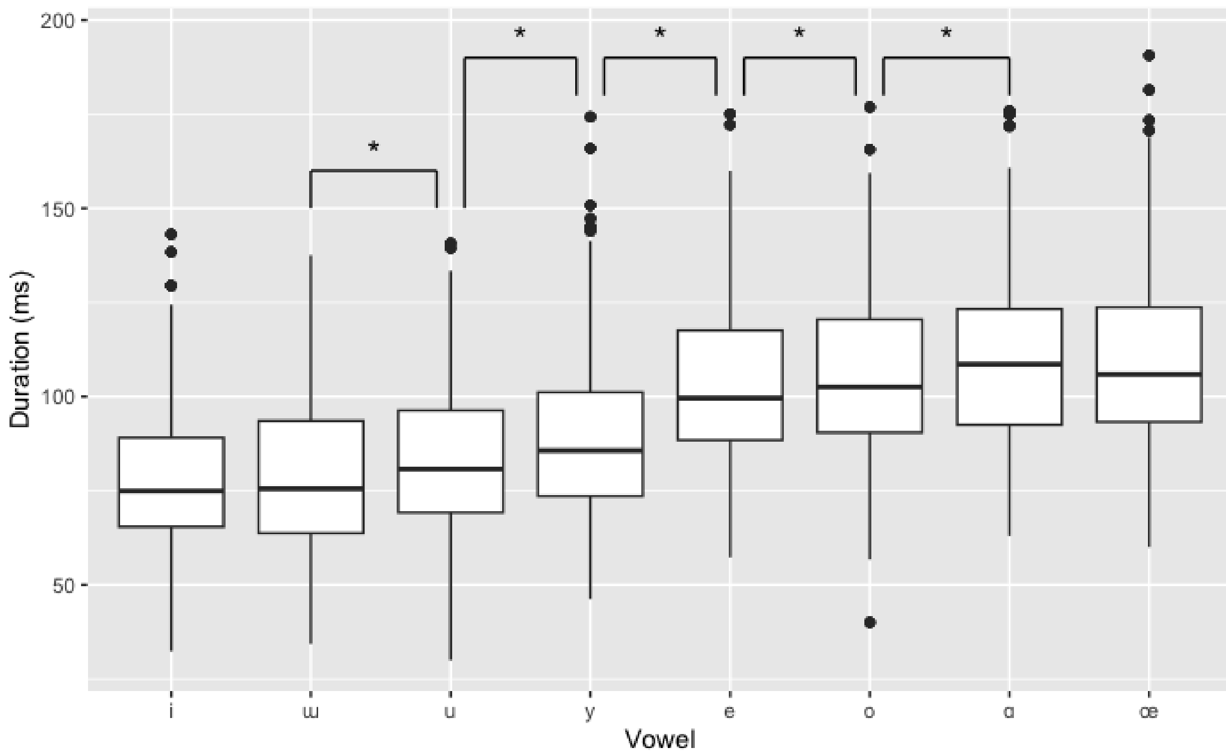

Figure 2.3. Vowel duration in monosyllabic nonwords. Asterisks indicate significant differences in duration between vowel pairs.

Table 2.2. Vowel duration (log-transformed) model results for vowels in monosyllabic nonwords.

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	4.50	0.04	26	107.42	< 0.0001***
/ɑ/ vs. /œ/	0.002	0.01	2051	0.15	0.88
/o/ vs. /ɑ/	0.04	0.01	2051	3.75	< 0.0001***
/e/ vs. /o/	0.04	0.01	2051	2.93	0.002**
/y/ vs. /e/	0.16	0.01	2051	16.79	< 0.0001***
/u/ vs. /y/	0.06	0.01	2051	5.79	< 0.0001***
/ɯ/ vs. /u/	0.06	0.01	2051	6.26	< 0.0001***
/i/ vs. /ɯ/	0.004	0.01	2051	0.37	0.71
Speaker gender	0.04	0.01	26	0.92	0.37

In the F1 model, there was a significant main effect of vowel category on vowel F1 ($F(7, 2051) = 2245.18$, $p < 0.0001$), indicating that vowel F1 varies across the vowel categories (Table 2.3). There were significant differences in vowel F1 between all vowel pairs (all *p*s < 0.0001) except for between /œ/ and /o/ ($p = 0.28$; Figure 2.4). The results here are in line with the description of canonical vowel height in Turkish in the literature whereby /i y ɯ u/ are phonetically high, /e œ o/ are phonetically mid, and /ɑ/ is phonetically lower than /e œ o/ (e.g., Göksel & Kerslake, 2005; Kopkalli-Yavuz, 2010).The highest differences in F1 were between the vowel pair /ɯ œ/ ($ß = -109.66$, $SE = 3.85$, $t = -28.47$, $p < 0.0001$) and the vowel pair /e ɑ/ ($ß = -140.90$, $SE = 3.85$, $t = -36.58$, $p < 0.0001$).

Table 2.3. Vowel F1 model results for vowels in monosyllabic nonwords.

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	480.50	6.25	26	76.85	< 0.0001***
/i/ vs. /y/	-26.62	3.85	2051	-6.92	< 0.0001***
/y/ vs. /u/	-25.12	3.85	2051	-6.52	< 0.0001***
/u/ vs. /ɯ/	-24.67	3.85	2051	-6.41	< 0.0001***
/ɯ/ vs. /œ/	-109.66	3.85	2051	-28.47	< 0.0001***
/œ/ vs. /o/	-4.13	3.85	2051	-1.07	0.28
/o/ vs. /e/	-49.17	3.85	2051	-12.77	< 0.0001***
/e/ vs. /ɑ/	-140.90	3.85	2051	-36.58	< 0.0001***
Speaker gender	47.87	6.25	26	7.50	< 0.0001***

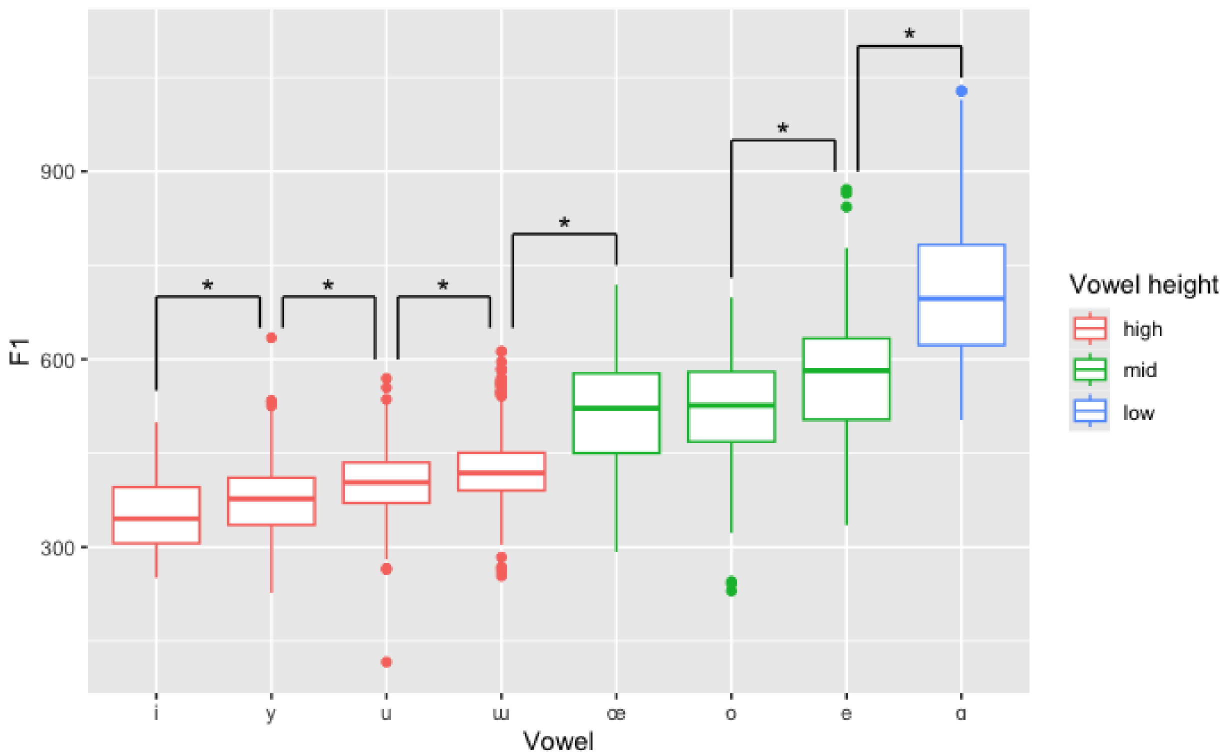

Figure 2.4. Vowel F1 in monosyllabic nonwords. Asterisks indicate significant differences in F1 between vowel pairs.

In the F2 model, there was a significant main effect of vowel category on vowel F2 ($F(7, 2051) = 3862.66$, $p < 0.0001$), indicating that vowel F2 varies across the vowel categories (Table 2.4). There were significant differences in vowel F2 between all vowel pairs (all *p*s < 0.0001; Figure 2.5). However, these differences are not necessarily such that phonologically front and phonologically back vowels are phonetically clearly differentiated in their F2. More specifically, there are larger differences in F2 between front vowels /œ-e/ and /y-i/ than between the vowels /ɯ-œ/ that are at the boundary of back and front vowels. The results here agree with Kopkalli-Yavuz (2010) that canonical Turkish vowels constitute a continuum along F2 rather than exhibit discontinuities in F2 between front and back vowels. Similar to Kopkalli-Yavuz, I find that /ɑ ɯ œ/ approach centrality, although this pattern toward centrality is less severe in my data with larger distance between /ɯ œ/ than between /ɑ ɯ/. Note that due to the low back unrounded vowel having lower F2 than the high back unrounded vowel /ɯ/, I prefer to transcribe it as /ɑ/ rather than as the low central unrounded vowel /a/. Lastly, in terms of the relationship between vowel rounding and vowel F2, the results here are in line with Kopkalli-Yavuz's argument that rounded vowels are more back than their unrounded counterparts within the same backness class, suggesting that F2 might be a cue to vowel roundess in addition to vowel backness.

Table 2.4. Vowel F2 model results for vowels in monosyllabic nonwords.

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	1402.66	9.91	26	141.55	< 0.0001***
/u/ vs. /o/	-41.48	11.00	2051	-3.77	0.0002***
/o/ vs. /ɑ/	-272.975	10.99	2051	-24.84	< 0.0001***
/ɑ/ vs. /ɯ/	-33.52	10.99	2051	-3.05	0.002**
/ɯ/ vs. /œ/	-192.73	10.99	2051	-17.52	< 0.0001***
/œ/ vs. /y/	-276.66	11.00	2051	-25.15	< 0.0001***
/y/ vs. /e/	-159.28	11.00	2051	-14.48	< 0.0001***
/e/ vs. /i/	-424.49	11.00	2051	-38.58	< 0.0001***
Speaker gender	95.32	9.91	26	9.62	< 0.0001***

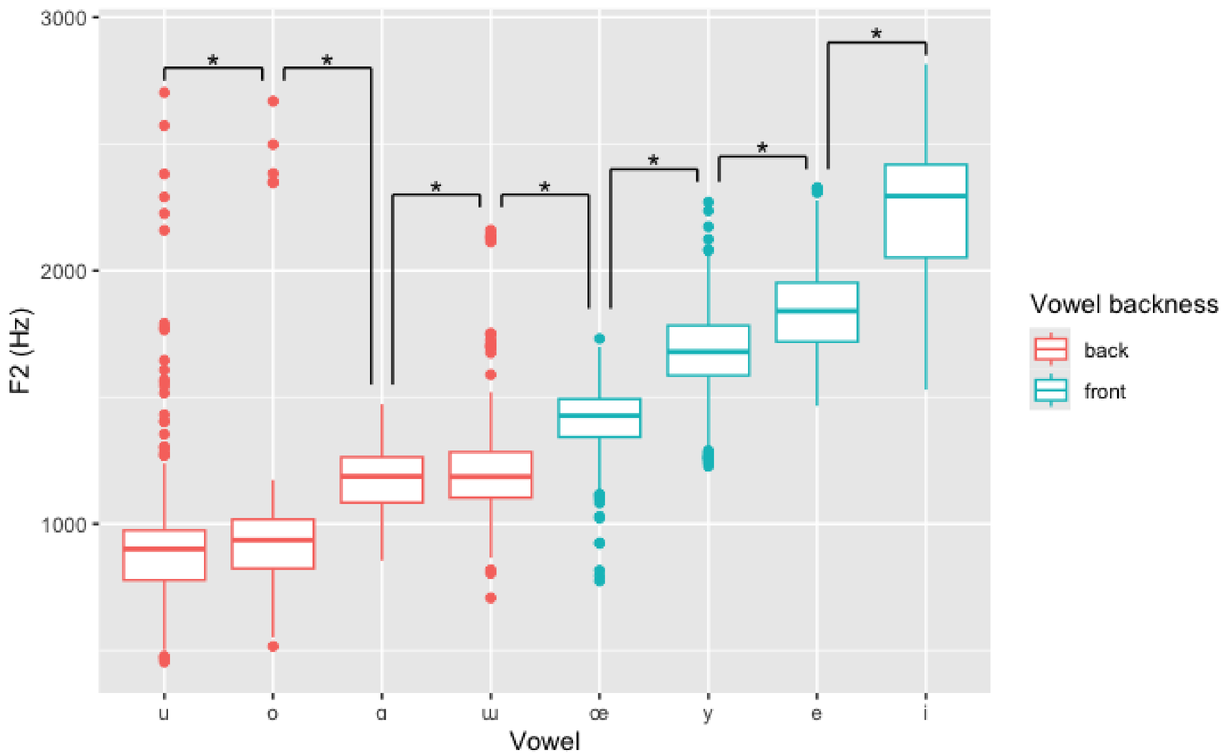

Figure 2.5. Vowel F2 in monosyllabic nonwords. Asterisks indicate significant differences in F2 between vowel pairs.

Overall, the results from monosyllabic nonwords in my data suggest that whereas canonical Turkish vowels are well separated in terms of vowel height with vowels being classified into three groups according to their F1, a more continuous distribution of vowels emerge for vowel backness based on F2. I argue that these results suggest that Turkish vowel phonology is not a mirror image of canonical Turkish vowel phonetics. In the upcoming sections, I present a phonetic analysis of canonical Turkish vowels in disyllabic nonwords to examine the effects of vowel position within a word and vowel-to-vowel coarticulation to examine whether vowels preserve their contrasts in more complex contexts that might increase phonetic variability.

2.2.2. Turkish vowels in disyllabic nonwords

There were 9980 vowel tokens in total that were produced in disyllabic /bVb.bVb/ nonwords (Figure 2.6). The phonetic measures of interest were vowel duration (log-transformed) and vowel F1 and F2 at vowel midpoint, which were modeled as the dependent variable using linear mixed effects models with vowel category, syllable number (V_1, V_2), speaker gender, and the interaction of vowel category and syllable number as predictors. There were random intercepts for speaker and item (5) . Multi-level fixed effects were evaluated via a type III ANOVA using Satterthwaite's Method. All predictors were sum coded (for syllable number, V_1: -1, V_2: 1). Planned comparisons between levels of a variable were performed using the `emmeans` package (Lenth & Piaskowski, 2025).

(5) vowel duration or formant ~ vowel category
+ syllable number
+ speaker gender
+ (1 | speaker)
+ (1 | item)

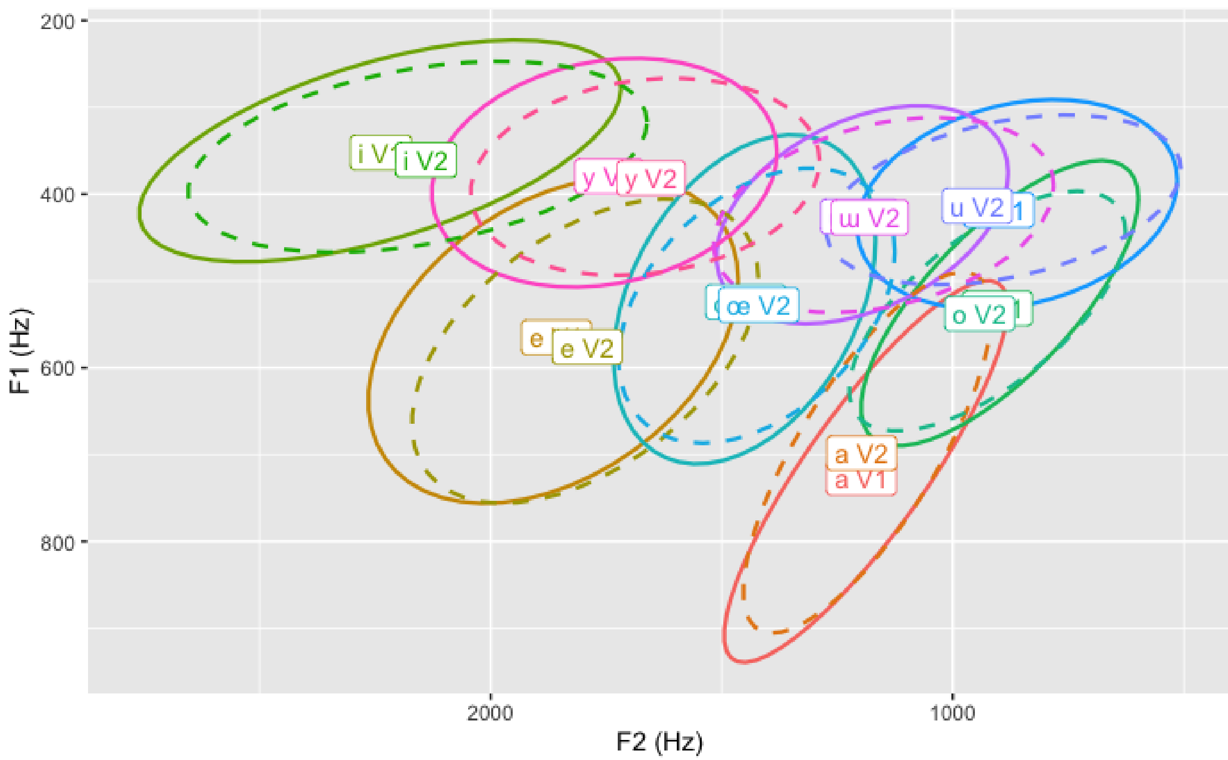

Figure 2.6. Vowel ellipses depicting the vowels produced in V_1 position (solid ellipses) and V_2 position (dashed ellipses) in the disyllabic nonwords, based on F1 and F2 measurements taken at vowel midpoint.

In the duration model, there was a significant main effect of vowel category on vowel duration ($F(7, 4329.1) = 306.17$, $p < 0.0001$), indicating that vowel duration varied across the vowel categories. There was a significant main effect of syllable number ($\beta = -0.02$, $SE = 0.002$, $t = -11.56$, $p < 0.0001$), indicating that vowels in the second syllable of the disyllabic nonwords were shorter by 2.68 ms on average than vowels in the first syllable of the disyllabic nonwords (Table 2.5). There was also a significant interaction of vowel category and syllable number ($F(7, 4329.7) = 3.89$, $p = 0.0003$). Planned comparisons revealed that all vowels were significantly shorter in the second syllable compared to the first syllable (ps < 0.05) with the exception of /e/ and /o/ (ps > 0.1; Figure 2.7).

Table 2.5. Vowel duration model results for disyllabic nonwords.

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	4.36	0.03	27.31	127.58	< 0.0001***
/ɑ/	0.13	0.01	4331.06	20.67	< 0.0001***
/e/	0.07	0.01	4331.06	11.22	< 0.0001***
/i/	-0.11	0.01	4331.06	-16.80	< 0.0001***
/o/	0.04	0.01	4331.06	-23.33	< 0.0001***
/œ/	0.17	0.01	4327.11	5.73	< 0.0001***
/u/	-0.11	0.01	4327.11	26.51	< 0.0001***
/ɯ/	-0.15	0.01	4323.16	-16.83	< 0.0001***
Syllable number	-0.02	0.00	9879.80	-11.56	< 0.0001***
Speaker gender	0.03	0.03	26.00	0.89	0.38
/ɑ/: Syllable number	0.01	0.01	4330.39	0.95	0.34
/e/: Syllable number	0.01	0.01	4330.39	1.85	0.06
/i/: Syllable number	-0.01	0.01	4330.39	-1.29	0.20
/o/: Syllable number	0.02	0.01	4330.39	2.58	0.01**
/œ/: Syllable number	0.004	0.01	4329.11	0.59	0.55
/u/: Syllable number	-0.02	0.01	4329.11	-2.76	0.006**
/ɯ/: Syllable number	0.007	0.01	4327.82	1.12	0.27

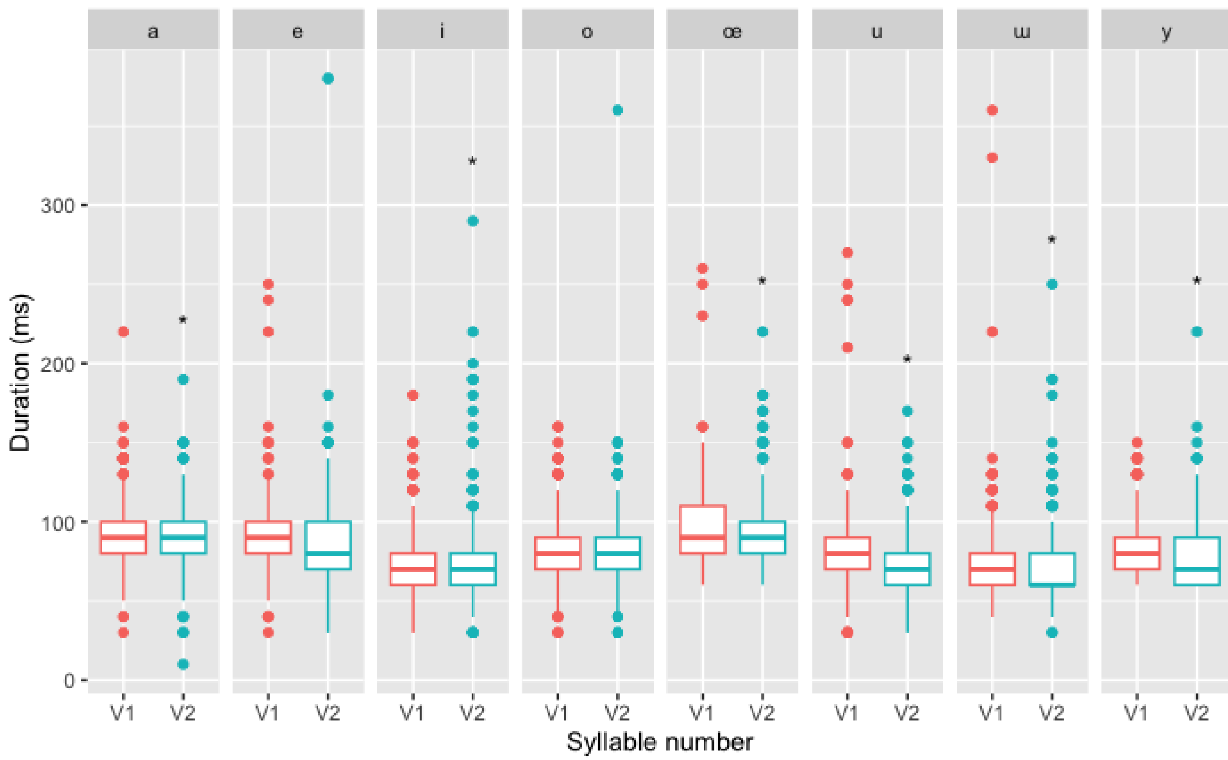

Figure 2.7. Vowel duration in V_1 and V_2 vowel position in disyllabic nonwords. Asterisks indicate significant differences in duration between vowel positions.

In the F1 model, there was a significant main effect of vowel category (F(7, 449.6.1) = 4414.20, $p < 0.0001$), indicating that vowel F1 varied across the vowel categories. There was no significant main effect of syllable number ($\beta = -0.08$, $SE = 0.54$, $t = -0.15$, $p = 0.88$), however, there was a significant interaction of vowel category and syllable number (F(7, 449.8) = 13.76, $p < 0.0001$), indicating that syllable number had an effect on vowel F1 that varies by vowel category (Table 2.6). Planned comparisons revealed that /ɑ/ vowels had significantly higher F1 in the first syllables of disyllabic nonwords than in the second syllables of disyllabic nonwords ($\Delta F1 = 30.14$, $SE = 3.49$, $z = 8.63$, $p < 0.0001$) whereas /e/ and /i/ vowels had significantly lower F1 in the first syllables of disyllabic nonwords than in the second syllables of disyllabic nonwords ($\Delta F1 = 10.02$, $SE = 3.49$, $z = -2.87$, $p = 0.004$; $\Delta F1 = 12.33$, $SE = 3.49$, $z = -3.53$, $p = 0.0004$, respectively). None of the other pairwise comparisons were significant (Figure 2.8).

Table 2.6. Vowel F1 results for disyllabic nonwords.

	Estimate	**Std. Error**	**df**	**t value**	**Pr(>|t|)**
(Intercept)	476.59	5.88	26.46	81.02	< 0.0001***
/ɑ/	221.04	1.66	449.75	133.21	< 0.0001***
/e/	78.77	1.66	449.75	47.47	< 0.0001***
/i/	-133.78	1.66	449.75	-80.62	< 0.0001***
/o/	43.79	1.66	449.54	26.38	< 0.0001***
/œ/	35.08	1.66	449.54	21.14	< 0.0001***
/u/	-73.47	1.66	449.32	-44.25	< 0.0001***
/ɯ/	-62.88	1.66	449.75	-37.90	< 0.0001***
Syllable number	-0.08	0.54	9873.07	-0.15	0.88
Speaker gender	44.79	5.86	26.00	7.65	< 0.0001***
/ɑ/: Syllable number	-14.99	1.66	449.60	-9.03	< 0.0001***
/e/: Syllable number	5.09	1.66	449.60	3.07	0.002**
/i/: Syllable number	6.24	1.66	449.60	3.76	< 0.0001***
/o/: Syllable number	2.70	1.66	449.99	1.63	0.10
/œ/: Syllable number	1.54	1.66	449.99	0.93	0.35
/u/: Syllable number	-1.79	1.66	450.38	-1.08	0.28
/ɯ/: Syllable number	0.74	1.66	449.60	0.45	0.66

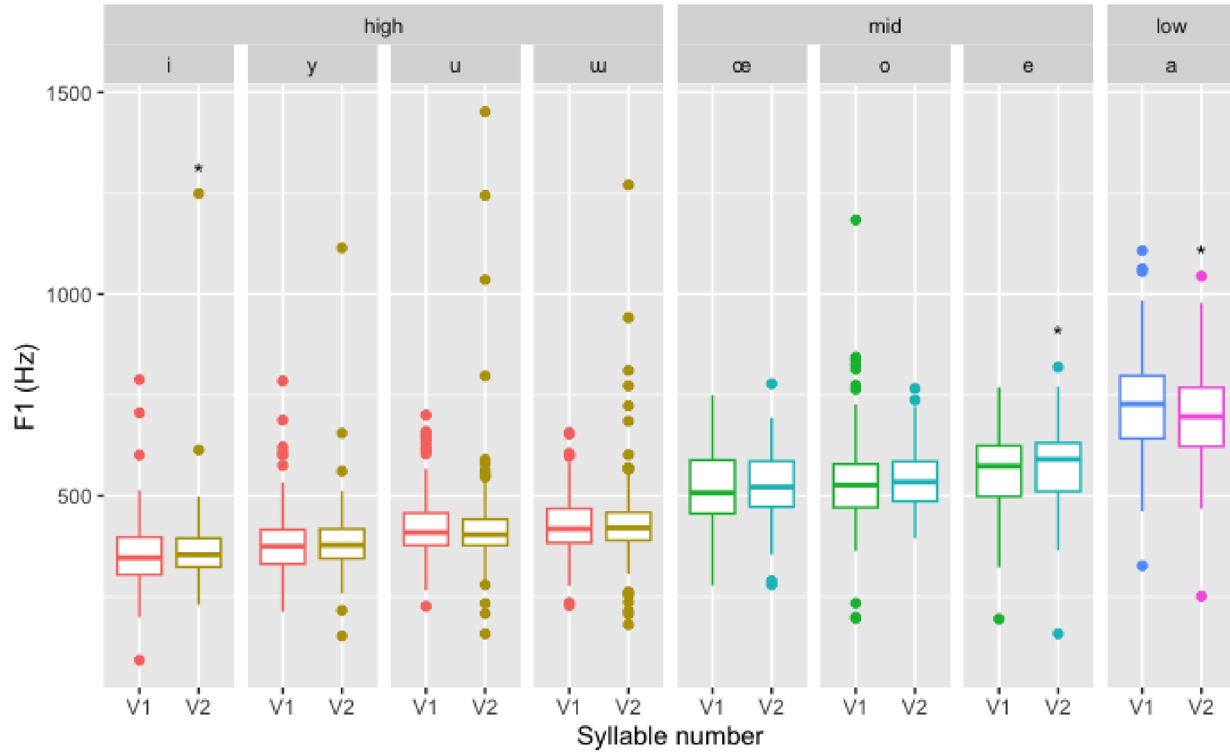

Figure 2.8. Vowel F1 in V_1 and V_2 vowel position in disyllabic nonwords. Asterisks indicate significant differences in duration between vowel positions.

In the F2 model, there was a significant main effect of vowel category ($F(7, 359.0) = 6260.95$, $p < 0.0001$), indicating that vowel F2 varied across the vowel categories. There was a significant main effect of syllable number ($\beta = -13.89$, $SE = 1.80$, $t = -7.71$, $p < 0.0001$), indicating that vowels in the second syllables of disyllabic nonwords had lower F2 than vowels in the first syllables of disyllabic nonwords. There was also a significant interaction of vowel category and syllable number ($F(7, 359.1) = 22.28$, $p < 0.0001$), indicating that syllable number had an effect on vowel F2 that varies by vowel category (Table 2.7). Planned comparisons revealed that /e i œ y/ vowels had significantly higher F2 in the first syllables of disyllabic nonwords than in the second syllables of disyllabic nonwords (/e/: $\Delta F2 = 63.56$, $SE = 11.2$, $z = 5.66$, $p < 0.0001$; /i/: $\Delta F2 = 96.9$, $SE = 11.2$, $z = 8.64$, $p < 0.0001$; /œ/: $\Delta F2 = 31.94$, $SE = 11.2$, $z = 2.84$, $p = 0.005$; /y/: $\Delta F2 = 91.48$, $SE = 11.2$, $z = 8.15$, $p < 0.0001$) whereas /o u/ vowels had significantly lower F2 in the first syllables of disyllabic nonwords than in the second syllables of disyllabic nonwords (*ΔF1* = 34.95, *SE* = 11.2, *z* = -3.11, *p* = 0.002; *ΔF1* = 18.60, *SE* = 11.2, *z* = --3.92, *p* = 0.0001, respectively). None of the other pairwise comparisons were significant (Figure 2.9).

Table 2.7. Vowel F2 results for disyllabic nonwords.

	Estimate	**Std. Error**	**df**	**t value**	**Pr(>|t|)**
(Intercept)	1395.75	10.92	26.91	127.77	< 0.0001***
/ɑ/	-223.40	5.32	359.01	-42.01	< 0.0001***
/e/	401.21	5.32	359.01	75.45	< 0.0001***
/i/	763.22	5.32	359.01	143.52	< 0.0001***
/o/	-497.73	5.32	358.93	-93.57	< 0.0001***
/œ/	13.26	5.32	358.93	2.49	0.01*
/u/	-498.05	5.32	358.84	-93.60	< 0.0001***
/ɯ/	-230.73	5.32	359.01	-43.39	< 0.0001***
Syllable number	-13.89	1.80	9877.39	-7.71	< 0.0001***
Speaker gender	86.37	10.83	26.00	7.98	< 0.0001***
/ɑ/: Syllable number	14.52	5.32	358.88	2.73	0.007**
/e/: Syllable number	-17.89	5.32	358.88	-3.36	< 0.0001***
/i/: Syllable number	-34.60	5.32	358.88	-6.51	< 0.0001***
/o/: Syllable number	31.37	5.32	359.32	5.90	< 0.0001***
/œ/: Syllable number	-2.08	5.32	359.32	-0.39	0.70
/u/: Syllable number	35.93	5.32	359.76	6.75	< 0.0001***
/ɯ/: Syllable number	4.59	5.32	358.88	0.87	0.39

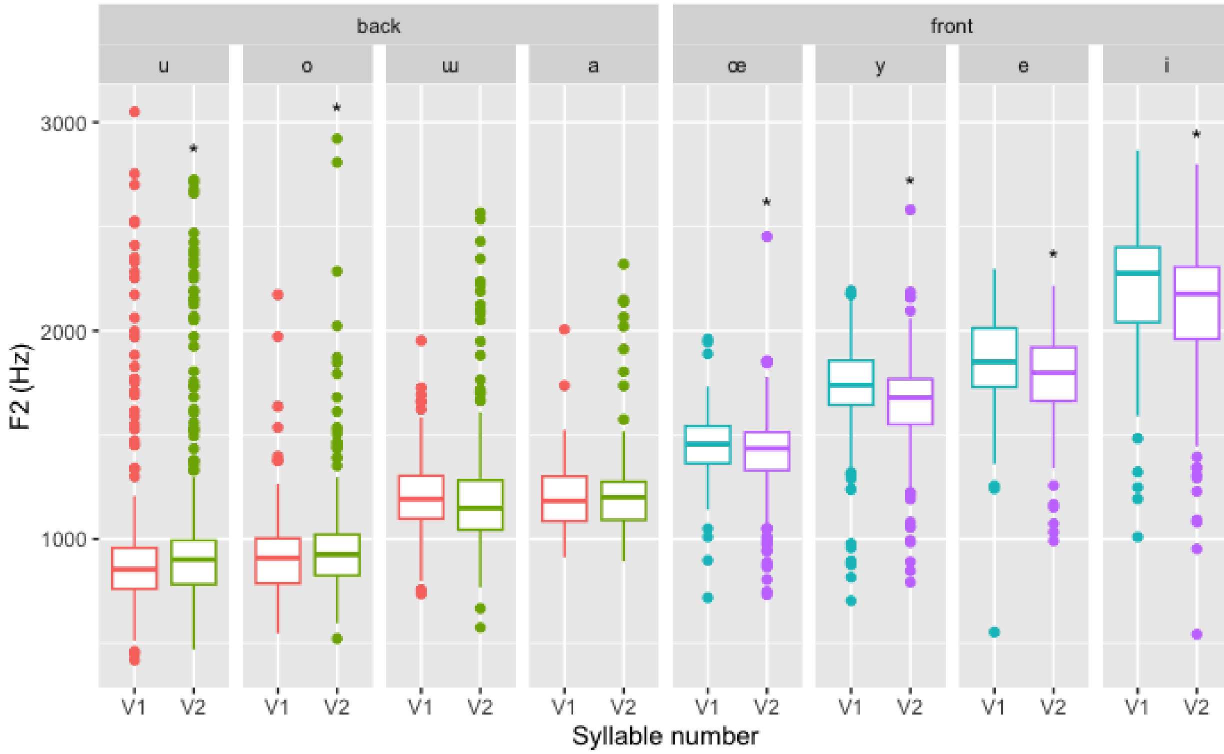

Figure 2.9. Vowel F2 in V_1 and V_2 vowel position in disyllabic nonwords. Asterisks indicate significant differences in duration between vowel positions.

The findings from the pairwise comparisons of vowel F1 and F2 across V_1 and V_2 positions are summarized in Table 2.8. The vowel F2 findings suggest that front vowels are fronter in the first syllable position whereas back vowels are backer in the first syllable position, although this backing effect was not observed in all back vowels. This difference in vowel F2 across the V_1 and V_2 positions might be related to the vowel duration differences observed across these positions, whereby longer vowels in V_1 position overall have more enhanced vowel backness and shorter vowels in V_2 position are more centralized. The fronting of front vowels and backing of back vowels in initial syllables are in line with accounts of hyperarticulation and contrast enhancement under prominence (e.g., Cho, 2005; Foster & Cole, 2024), suggesting that word-initial syllables might be phonologically more prominent in Turkish (Barnes, 2002) to facilitate the accurate perception of vowels in non-initial syllables via vowel harmony (cf. Suomi, 1983; see Chapter 1).

Table 2.8. Summary of the syllable number effects on vowel F1 and F2 in the pairwise comparisons.

Vowel	Height	F1 in V_1 position	Backness	F2 in V_1 position
/i/	High	Higher	Front	Fronter
/y/	High		Front	Fronter
/ɯ/	High		Back	
/u/	High		Back	Backer
/e/	Non-high (mid)	Higher	Front	Fronter
/œ/	Non-high (mid)		Front	Fronter
/o/	Non-high (mid)		Back	Backer
/ɑ/	Non-high (low)	Lower	Back	

Barnes (2002) argues that in Turkish, the vowel in the word-initial syllable carries prominence due to the role it plays in setting the backness features of the following vowels in the word, and this prominence manifests in longer vowel durations. The findings here extend Barnes's finding and suggest that in addition to longer duration, vowels in the word-initial syllable might have enhanced F2. However, this pattern of phonological enhancement of the vowel's backness feature in the word-initial syllable is incomplete as no significant change in vowel F2 is observed for the back vowels /ɑ ɯ/. These two vowels are argued to be phonetically approaching centrality as described above. Hence, enhancement of vowel F2 is either phonetic with no enhancement for the more central back vowels, or interacts with vowel rounding whereby only rounded vowels exhibit backing in the word-initial syllable.

The small number of significant pairwise comparisons of vowel F1 across V_1 and V_2 positions is not readily explained by the contrast enhancement account proposed above. Only 3 out of 8 Turkish vowels exhibit significant effects of vowel position on F1: /ɑ e i/. Of the vowels that show an effect of vowel position on F1, the low vowel /ɑ/ is lower in V_1 position and the high vowel /i/ is higher in V_1 position. As such, /ɑ/ and /i/ behave in line with contrast enhancement in V_1 position and vowel reduction in V_2 position. In addition to /ɑ/ and /i/, the mid vowel /e/ is also significantly higher in V_1 position in my data. This change in F1 cannot be attributed to contrast enhancement because /e/ is a phonetically mid and phonologically nonhigh vowel: based on its phonetic height, /e/ would be predicted to not exhibit changes in vowel F1, and based on its phonological height, /e/ would be predicted to lower under phonological contrast enhancement. I argue that /e/ raising in V_1 position might be motivated by contrast preservation, as /e/ lowering to [ɛ] in non-initial syllables is an allophonic variation that has been proposed in the literature previously (Ergenç & Uzun, 2020; see below).

Overall, these results suggest that Turkish vowels might be more resistant to movements in F1 compared to movements in F2, and that central-back vowels are also resistant to movements in F2. I argue that these patterns of resistance are motivated by contrast preservation, as the vowels that exhibit formant movements are those that are more distant from the neighboring vowels in the vowel space. In the upcoming sections, I first compare vowel F1 and F2 across monosyllabic and disyllabic nonwords and then analyze anticipatory and carryover vowel-to-vowel coarticulation effects on F1 and F2 in disyllabic nonwords to examine coarticulatory vowel formant movements.

2.2.3. Turkish vowels by position in monosyllabic and disyllabic words

To examine whether vowels are realized differently based on their position in the word, I compared vowel formants from vowels in monosyllabic nonwords with vowels in disyllabic nonwords in which the first and second vowels are phonemically identical. The phonetic measures of interest were vowel F1 and vowel F2 at vowel midpoint. Each of these measures were modeled as the dependent variable in separate linear mixed effects models with vowel category, vowel position, speaker gender, and the interaction of vowel category and vowel position as predictors. There were random intercepts for speaker but not word due to singularity (6). Vowel category was sum coded. Vowel position had three levels: vowel in monosyllabic nonword (baseline), first vowel in disyllabic nonword (V_1), second vowel in disyllabic nonword (V_2). Fixed effects with more than 2 levels were evaluated via type III ANOVA using Satterthwaite's Method. Planned comparisons between levels of a variable were performed using the `emmeans` package (Lenth & Piaskowski, 2025) and Bonferroni correction.

(6) vowel formant ~ vowel category
+ vowel position
+ speaker gender
+ vowel category : vowel position
+ (1 | speaker)

In the F1 model, there was a significant simple effect of vowel category at the monosyllabic nonword level ($F(7, 3220) = 2182.74$, $p < 0.0001$) and a significant simple effect of speaker gender at the monosyllabic nonword level ($\beta = 46.76$, $SE = 5.99$, $t = 7.80$, $p < 0.0001$), indicating that vowel category and speaker gender significantly influenced vowel F1 in monosyllabic nonwords. There was no significant main effect of vowel position ($F(2, 3223.3) = 1.18$, $p = 0.31$), indicating that vowel position did not consistently influence vowel F1 across vowel categories. Neither V_1 nor V_2 vowels in disyllabic words had a significant effect on vowel F1 compared to vowels in monosyllabic nonwords (V_1: $\beta = 0.95$, $SE = 2.11$, $t = 0.45$, $p = 0.45$; V_2: $\beta = -2.78$, $SE = 2.11$, $t = -1.31$, $p = 0.19$). There was a significant interaction of vowel category and vowel position ($F(14, 3220) = 4.93$, $p = 3.759e-09$), indicating that the effect of vowel position on vowel F1 differed by vowel category (Table 2.9).

Table 2.9. F1 model results.

	Estimate	SE	t	p
(Intercept)	480.60	6.03	79.68	< .0001***
/ɑ/	238.21	2.71	88.03	< .0001***
/e/	97.39	2.71	35.92	< .0001***
/i/	-144.77	2.71	-53.50	< .0001***
/o/	46.52	2.71	17.19	< .0001***
/œ/	43.96	2.71	16.24	< .0001***
/u/	-93.96	2.71	-34.66	< .0001***
/ɯ/	-68.85	2.71	-25.44	< .0001***
V1 (disyllabic)	0.95	2.11	0.45	.654
V2 (disyllabic)	-2.78	2.11	-1.31	.189
Speaker gender	46.76	5.99	7.80	< .0001***
/ɑ/: V1 (disyllabic)	-12.89	5.55	-2.32	.020*
/e/: V1 (disyllabic)	-7.59	5.55	-1.37	.172
/i/: V1 (disyllabic)	-3.06	5.55	-0.55	.582
/o/: V1 (disyllabic)	-8.50	5.55	-1.53	.126
/œ/: V1 (disyllabic)	-4.51	5.55	-0.81	.416
/u/: V1 (disyllabic)	22.32	5.55	4.02	< .0001***
/ɯ/: V1 (disyllabic)	6.00	5.55	1.08	.280
/ɑ/: V2 (disyllabic)	-32.27	5.55	-5.82	< .0001***
/e/: V2 (disyllabic)	-3.79	5.55	-0.68	.495
/i/: V2 (disyllabic)	17.33	5.55	3.12	.002**
/o/: V2 (disyllabic)	-0.89	5.55	-0.16	.872
/œ/: V2 (disyllabic)	-9.36	5.55	-1.69	.092
/u/: V2 (disyllabic)	16.22	5.55	2.92	.004**
/ɯ/: V2 (disyllabic)	6.68	5.55	1.20	.229

Planned comparisons revealed that several vowel categories had significantly different F1 in the various vowel positions (Figure 2.10). For the vowel /ɑ/, vowels in V_2 position in disyllabic words had significantly lower F1 compared to vowels in monosyllabic words ($\Delta F1$ = 34.04 Hz, SE = 5.94, z = 5.90, p < 0.0001) and vowels in V_1 position in disyllabic words ($\Delta F1$ = 23.10 Hz, SE = 7.32, z = 3.15, p = 0.005). For the vowel /i/, vowels in V_2 position in disyllabic words had significantly higher F1 compared to vowels in monosyllabic words ($\Delta F1$ = -14.56 Hz, SE = 5.94, z = -2.45, p = 0.04). For the vowel /u/, vowels in V_1 position in disyllabic words had significantly higher F1 compared to vowels in monosyllabic words (*ΔF1* = -23.27 Hz, *SE* = 5.94, *z* = -3.92, *p* = 0.0003).

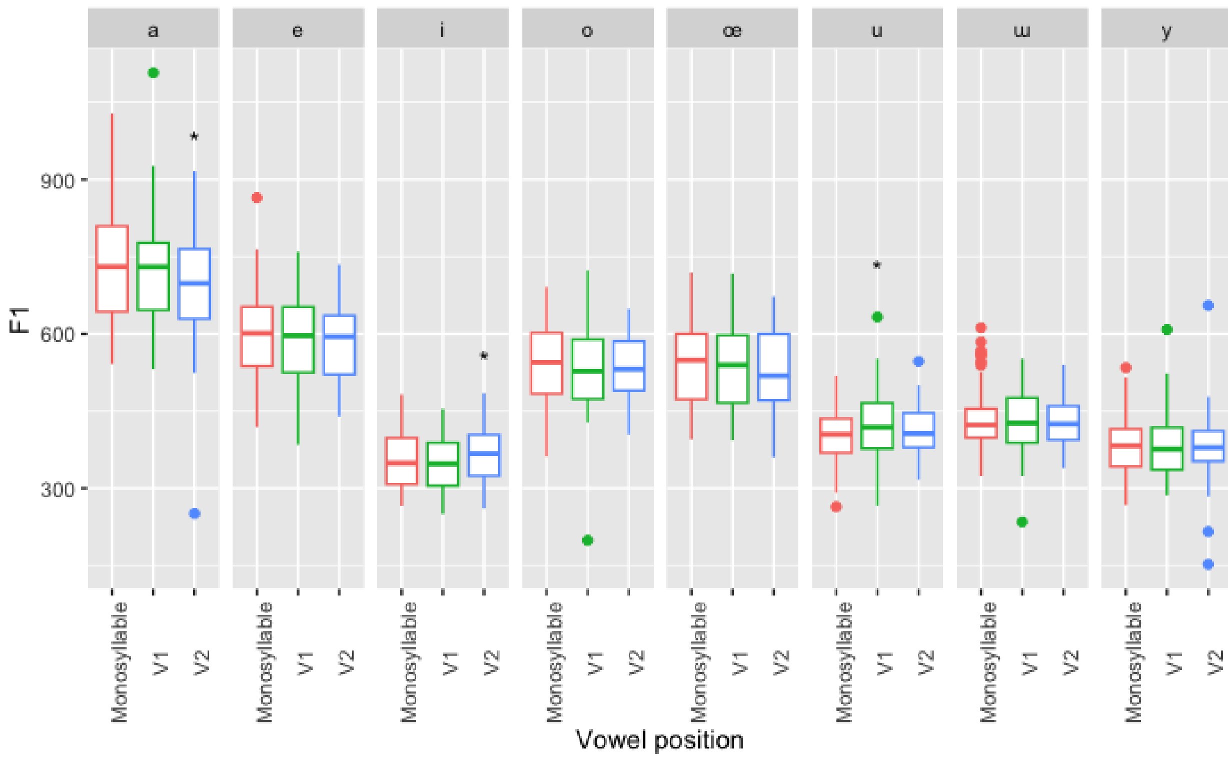

Figure 2.10. Boxplots showing F1 at vowel midpoint by vowel position in monosyllabic (Monosyllable) and disyllabic words (V1 or V2), faceted by vowel category. Asterisks indicate significant pairwise, Bonferroni-corrected comparisons between vowel F1 in monosyllabic nonwords and vowel F2 in V_1 or V_2 position of disyllabic nonwords.

In the F2 model, there was a significant simple effect of vowel category at the monosyllabic nonword level ($F(7, 3220) = 3823.96$, $p < 0.0001$) and a significant simple effect of speaker gender ($\beta = 91.88$, $SE = 10.04$, $t = 9.15$, $p < 0.0001$), indicating that vowel category and speaker gender significantly influenced vowel F2. There was also a significant main effect of vowel position ($F(2, 3227.6) = 16.44$, $p < 0.0001$), indicating that F2 values differed across positions. Compared to vowels in monosyllabic words, vowels in V1 position in disyllabic words did not have significantly different vowel F2 ($\beta = -2.21$, $SE = 6.04$, $t = -0.53$, $p = 0.59$) whereas vowels in V2 position in disyllabic words had significantly lower vowel F2 across all vowel categories (*β* = -34.19, *SE* = 6.04, *t* = -5.66, *p* < 0.0001). A significant interaction between vowel category and vowel position (*F*(14, 3220) = 6.08, *p* < 0.0001) indicated that the effect of vowel position on F2 varied by vowel category (Table 2.10).

Table 2.10. F2 model results.

	Estimate	SE	t	p
(Intercept)	1403.19	10.22	137.28	< .0001***
/ɑ/	-240.44	7.74	-31.07	< .0001***
/e/	423.88	7.75	54.69	< .0001***
/i/	849.52	7.74	109.79	< .0001***
/o/	-514.85	7.74	-66.54	< .0001***
/œ/	-16.56	7.74	-2.14	.032*
/u/	-558.87	7.75	-72.10	< .0001***
/ɯ/	-207.06	7.74	-26.76	< .0001***
V1 (disyllabic)	-3.21	6.04	-0.53	.594
V2 (disyllabic)	-34.19	6.04	-5.66	< .0001***
Speaker gender	91.88	10.04	9.15	< .0001***
/ɑ/: V1 (disyllabic)	-0.45	15.87	-0.03	.977
/e/: V1 (disyllabic)	-40.09	15.87	-2.53	.012*
/i/: V1 (disyllabic)	-37.48	15.87	-2.36	.018*
/o/: V1 (disyllabic)	-7.45	15.87	-0.47	.639
/œ/: V1 (disyllabic)	26.88	15.87	1.69	.090
/u/: V1 (disyllabic)	29.72	15.87	1.87	.061
/ɯ/: V1 (disyllabic)	-2.40	15.87	-0.15	.880
/ɑ/: V2 (disyllabic)	17.38	15.87	1.10	.273
/e/: V2 (disyllabic)	-22.41	15.87	-1.41	.158
/i/: V2 (disyllabic)	-103.03	15.87	-6.49	< .0001***
/o/: V2 (disyllabic)	45.56	15.87	2.87	.004**
/œ/: V2 (disyllabic)	44.63	15.87	2.81	.005**
/u/: V2 (disyllabic)	62.86	15.87	3.96	< .0001***
/ɯ/: V2 (disyllabic)	-14.26	15.87	-0.90	.369

Planned comparisons revealed that several vowel categories had significantly different F2 values across positions (Figure 2.11). For the vowel /e/, vowels in V_1 and V_2 positions in disyllabic words had significantly lower F2 compared to vowels in monosyllabic words (V_1: 43.31 Hz, *SE* = 17.0, *z* = 2.55, *p* = 0.03; V_2: 56.60 Hz, *SE* = 17.0, *z* = 3.33, *p* = 0.003). For the vowel /i/, vowels in V_1 and V_2 positions had significantly lower F2 compared to monosyllabic words (V_1: 40.69 Hz, *SE* = 17.0, *z* = 2.40, *p* = 0.0496; V_2: 137.22 Hz, *SE* = 17.0, *z* = 8.08, *p* < 0.0001), and in disyllabic words, F2 was significantly lower in V_2 position compared to V_1 position ($\Delta F2$ = 96.53 Hz, *SE* = 20.9, *z* = 4.61, *p* < 0.0001). For the vowel /ɯ/, vowels in V_2 position had significantly lower F2 than monosyllabic vowels ($\Delta F2$ = 48.45 Hz, *SE* = 17.0, *z* = 2.85, *p* = 0.01). For the vowel /y/, vowels in V_2 position also had significantly lower F2 than monosyllabic vowels ($\Delta F2$ = 64.92 Hz, *SE* = 17.0, *z* = 3.82, *p* = 0.0004), and F2 was significantly lower in V_2 than V_1 ($\Delta F2$ = 92.96 Hz, *SE* = 20.9, *z* = 4.44, *p* < 0.0001). Finally, for the vowels /ɑ o u œ/, there were no significant differences in F2 across positions.

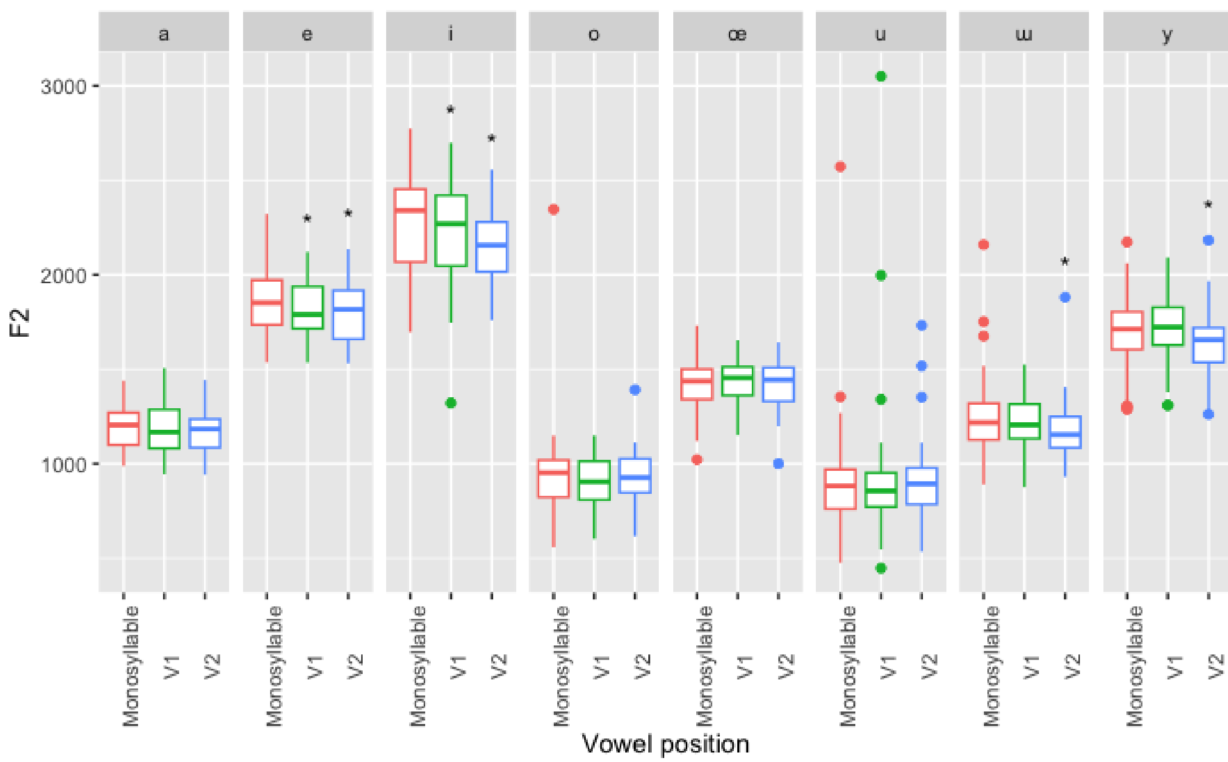

Figure 2.11. Boxplots showing F2 at vowel midpoint by vowel position in monosyllabic (Monosyllable) and disyllabic words (V1 or V2), faceted by vowel category. Asterisks indicate significant pairwise, Bonferroni-corrected comparisons between vowel F2 in monosyllabic nonwords and vowel F2 in V_1 or V_2 position of disyllabic nonwords.

Overall, these results show that whereas vowel position in word does not impact vowel F1 in a consistent way across vowels, vowel F2 is lower in second syllables of disyllabic words compared to vowels in monosyllabic words. However, pairwise comparisons revealed that 3 of the 4 back vowels did not have F2 differences across positions, whereas 3 of the 4 front vowels had lower F2 in second syllables of disyllabic words. In contrast, compared to vowel F1 in monosyllabic words, /ɑ/ had lower F1 whereas /i/ had higher F1 in second syllables of disyllabic words. Overall, these results suggest that vowels might be produced with more centralized articulations in noninitial syllables of multisyllabic words compared to monosyllabic words, although this centralization is observed in fewer number of vowels compared to the models in the previous section that examined the effects of vowel position in disyllabic nonwords only.

2.3. Vowel-to-Vowel Coarticulation

Previous research in vowel-to-vowel coarticulation has found that vowel space density is inversely related to degree of vowel-to-vowel coarticulation, arguing that lower degree of coarticulation in denser vowel spaces help preserve vowel contrasts (Manuel, 1990). However, this finding was not replicated by Beddor et al. (2002), suggesting that patterns of vowel-to-vowel coarticulation might be language specific. Beddor and Kopkallı-Yavuz (1995) measured vowel-to-vowel coarticulation in Turkish to examine whether as in Turkish vowel harmony, coarticulation is predominantly left-to-right, with carryover coarticulation being more robust than anticipatory coarticulation. In that study, 3 native Turkish speakers produced Turkish disyllabic words that make 38 minimal pairs with respect to whether or not a vowel is subject to vowel-to-vowel coarticulation (e.g., *deva* ‘remedy’ and *deve* ‘camel’ for the V_1 vowel /e/). The target and coarticulatory vowels in the stimuli were limited to /ɑ e i/ and the phonetic measure of interest was vowel F2, which was analyzed as difference scores within the minimal pairs. The results supported anticipatory (right-to-left) coarticulation in F2 whereas for carryover (left-to-right) coarticulation, only /ɑ/ exhibited coarticulatory shifts in F2. Where there was evidence in support of coarticulatory effects, shifts in F2 were greater at vowel offset for anticipatory coarticulation and at vowel onset for carryover coarticulation. Overall, the patterns of coarticulation varied between anticipatory and carryover coarticulation, and the authors suggest that word-final stress in Turkish might explain the greater tendency for vowels to undergo anticipatory coarticulation (see also Inkelas et al., 2001). In this section, I investigate how all 8 Turkish canonical vowels behave in all possible vowel-to-vowel coarticulation contexts by analyzing vowel F1 and F2 in disyllabic /bV_1b.bV_2b/ nonwords.

2.3.1. Anticipatory Coarticulation

There were 4990 vowel tokens in total that were produced in the first syllables of disyllabic nonwords. The phonetic measures of interest were vowel F1 and F2 at vowel offset, which were modeled as the dependent variable using linear mixed effects models with the interaction of V_1 vowel category and V_2 vowel category as predictors. There were random intercepts for speaker but not item due to singularity (7). Multi-level fixed effects were evaluated via a type III ANOVA using Satterthwaite's Method. Predictors were sum coded. Planned comparisons between levels of a variable were performed using the `emmeans` package (Lenth & Piaskowski, 2025) with Bonferroni adjustment for multiple comparisons.

(7) vowel formant ~ V_1 vowel category : V_2 vowel category
+ (1 | speaker)

In the F1 model, there was a significant interaction of V_1 vowel category and V_2 vowel category ($F(63, 4964) = 272.93$, $p < 0.0001$), indicating that vowels exhibited shifts in F1 that varied by target vowel category and context vowel category. Planned comparisons revealed several significant anticipatory coarticulation effects in 8 of the pairwise comparisons out of 64

(Table 2.11). These significant shifts in vowel F1 at V_1 vowel offset are in line with the predictions of anticipatory coarticulation whereby relatively lower F1 in V_2 caused V_1 to raise and relatively higher F1 in V_2 caused V_1 to lower, with the exception of V_1 /ɑ/ lowering in the context of relatively higher V_2 /e/ (Figure 2.12).

Table 2.11. Significant pairwise comparisons in the anticipatory coarticulation in vowel F1 model.

V_1 Vowel (target)	V_2 Vowel (context)	Estimate (ΔF1)	SE	z	p
/ɑ/	/e/	17.62	5.56	3.17	0.01*
/e/	/ɑ/	23.90	5.56	4.30	0.0001***
	/e/	21.41	5.56	3.85	0.001**
	/o/	22.07	5.56	3.97	0.001**
	/i/	-42.47	5.56	-7.64	< 0.0001***
	/y/	-24.73	5.56	-4.45	0.0001***
/o/	/ɯ/	-16.08	5.56	-2.89	0.03*
/y/	/u/	16.45	5.56	2.96	0.03*

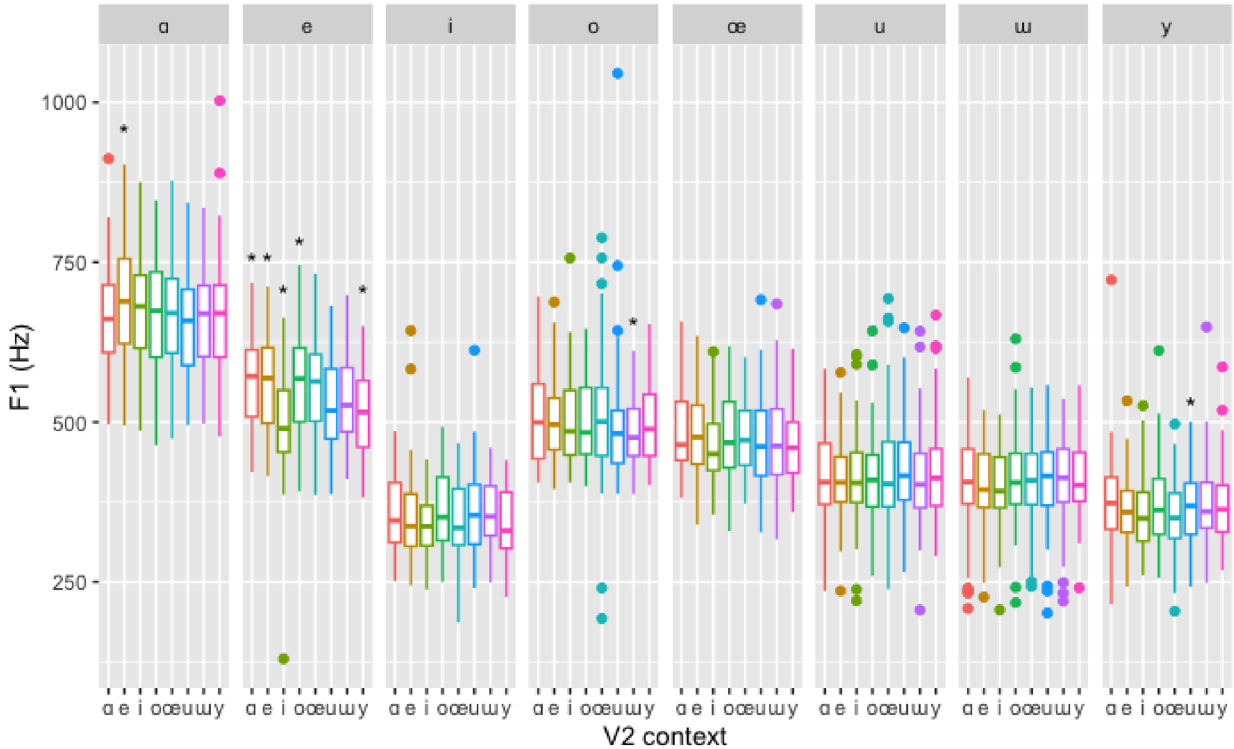

Figure 2.12. Vowel F1 at vowel offset for V_1 vowels. Asterisks indicate significant differences from the V_1 vowel mean F1.

In the F2 model, there was a significant interaction of V_1 vowel category and V_2 vowel category ($F(63, 4964) = 504{,}28$, $p < 0.0001$), indicating that vowels exhibited shifts in F2 that varied by target vowel category and context vowel category. Planned comparisons revealed several significant anticipatory coarticulation effects in 5 of the pairwise comparisons out of 64 (Table 2.12). These significant shifts in vowel F2 at V_1 vowel offset are in line with the predictions of anticipatory coarticulation whereby relatively lower F2 in V_2 caused V_1 to back and relatively higher F2 in V_2 caused V_1 to front (Figure 2.13).

Table 2.12. Significant pairwise comparisons in the anticipatory coarticulation in vowel F2 model.

V_1 Vowel (target)	V_2 Vowel (context)	Estimate (ΔF2)	SE	z	p
/ɑ/	/i/	48.62	17	2.87	0.03*
/e/	/i/	62.96	17	3.71	0.002**
	/œ/	-47.89	17	-2.87	0.04*
/u/	/o/	-59.87	17	-3.53	0.03*
/y/	/i/	61.56	17	3.63	0.002**

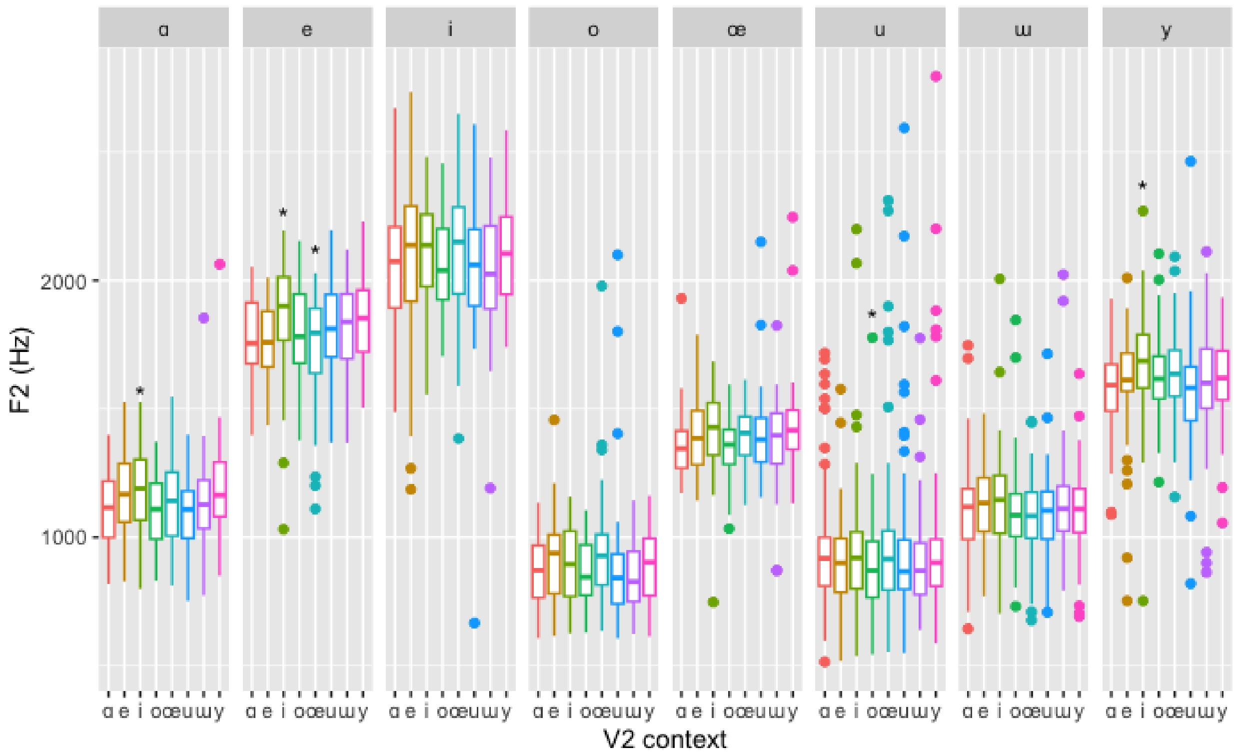

Figure 2.13. Vowel F2 at vowel offset for V_1 vowels. Asterisks indicate significant differences from the V_1 vowel mean F2.

Overall, these results support anticipatory coarticulation effects in Turkish vowels (Figure 2.14). Where there was a significant effect, in general, high V_2 vowels caused V_1 vowels to raise, non-high V_2 vowels caused V_1 vowels to lower, front V_2 vowels caused V_1 vowels to front, and back V_2 vowels caused V_1 vowels to back. However, these significant formant shifts are observed only for a limited number of cases and do not generalize over vowel feature classes. Most of the significant effects involved peripheral vowels /ɑ e i o u/ as targets and triggers of anticipatory coarticulation. In contrast, nonperipheral vowels /ɯ œ y/ did not exhibit or cause anticipatory coarticulation in most cases.

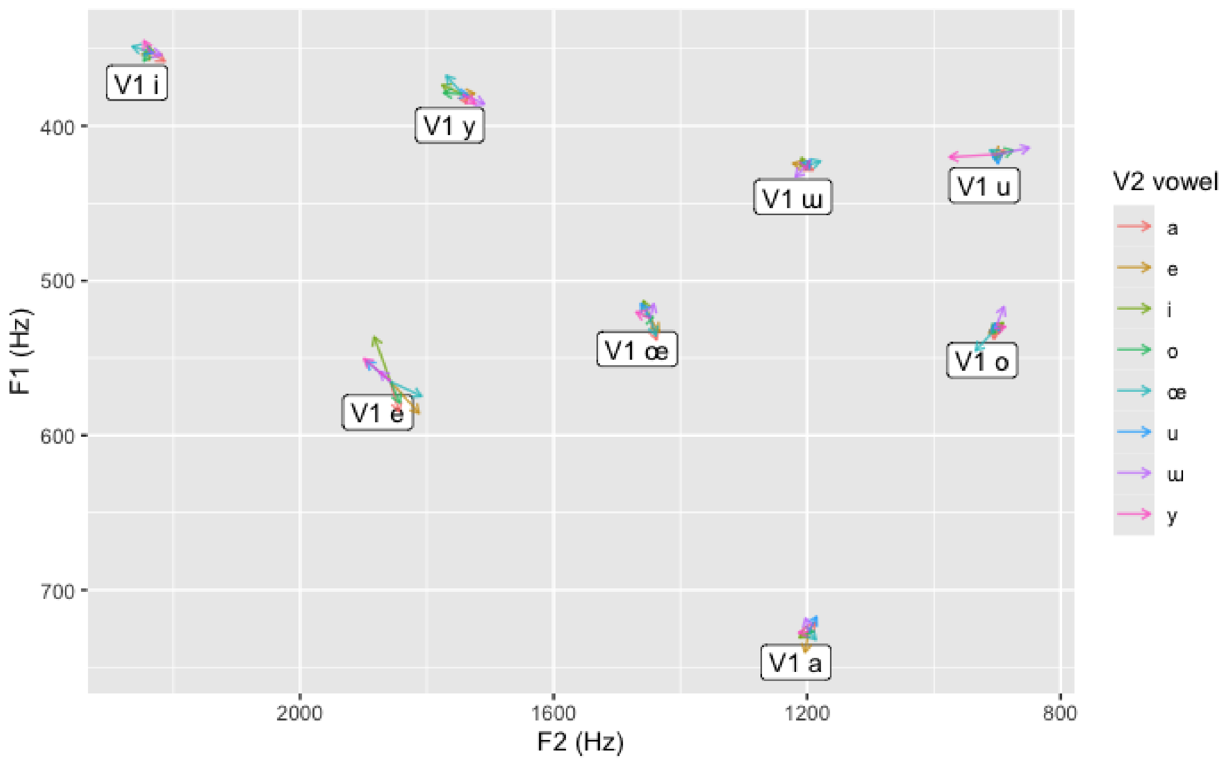

Figure 2.14. Anticipatory coarticulation effects on V_1 vowels by V_2 vowel context in disyllabic nonwords.

2.3.2. Carryover Coarticulation

There were 4990 vowel tokens in total that were produced in the second syllables of disyllabic nonwords. The phonetic measures of interest were vowel F1 and F2 at vowel onset, which were modeled as the dependent variable using linear mixed effects models with the interaction of V_1 vowel category and V_2 vowel category as predictors. There were random intercepts for speaker but not item due to singularity (8). Multi-level fixed effects were evaluated via a type III ANOVA using Satterthwaite's Method. Predictors were sum coded. Planned comparisons between levels of a variable were performed using the `emmeans` package (Lenth & Piaskowski, 2025) with Bonferroni adjustment for multiple comparisons.

(8) vowel formant ~ V_1 vowel category : V_2 vowel category
+ (1 | speaker)

In the F1 model, there was a significant interaction of V_1 vowel category and V_2 vowel category ($F(63, 4964) = 305.10$, $p < 0.0001$), indicating that vowels exhibited shifts in F1 that varied by target vowel category and context vowel category. Planned comparisons revealed several significant coarticulation effects (Figure 2.15). These significant shifts in vowel F1 at V_2 vowel onset are in line with the predictions of carryover coarticulation based on V_1 and V_2 vowel F1. Planned comparisons revealed several significant anticipatory coarticulation effects in 10 of the pairwise comparisons out of 64 (Table 2.13). These significant shifts in vowel F1 at V_2 vowel onset are generally in line with the predictions of carryover coarticulation whereby lower F1 in V_1 caused V_2 to raise and lower F1 in V_1 caused V_2 to lower, with the exception of V_2 /ɑ/ lowering in the context of V_1 /e/ and V_2 /u/ raising in the context of V_1 /ɯ/.

Table 2.13. Significant pairwise comparisons in the carryover coarticulation in vowel F1 model.

V_2 Vowel (target)	V_1 Vowel (context)	Estimate (ΔF1)	SE	z	p
/ɑ/	/ɑ/	25.71	5.22	4.93	< 0.0001***
	/e/	17.22	5.22	3.3	0.01**
	/y/	-24.06	5.22	-4.61	< 0.0001***
/e/	/y/	-28.24	5.22	-5.41	< 0.0001***
/i/	/o/	15.86	5.22	3.04	0.02*
/ɯ/	/ɑ/	21.86	5.22	4.19	0.0002**
	/y/	-17.78	5.22	-3.41	0.005**
/o/	/e/	16.67	5.22	3.2	0.01*
/œ/	/y/	-20.22	5.22	-3.88	0.001**
/u/	/ɯ/	-16.48	5.22	-3.16	0.01*

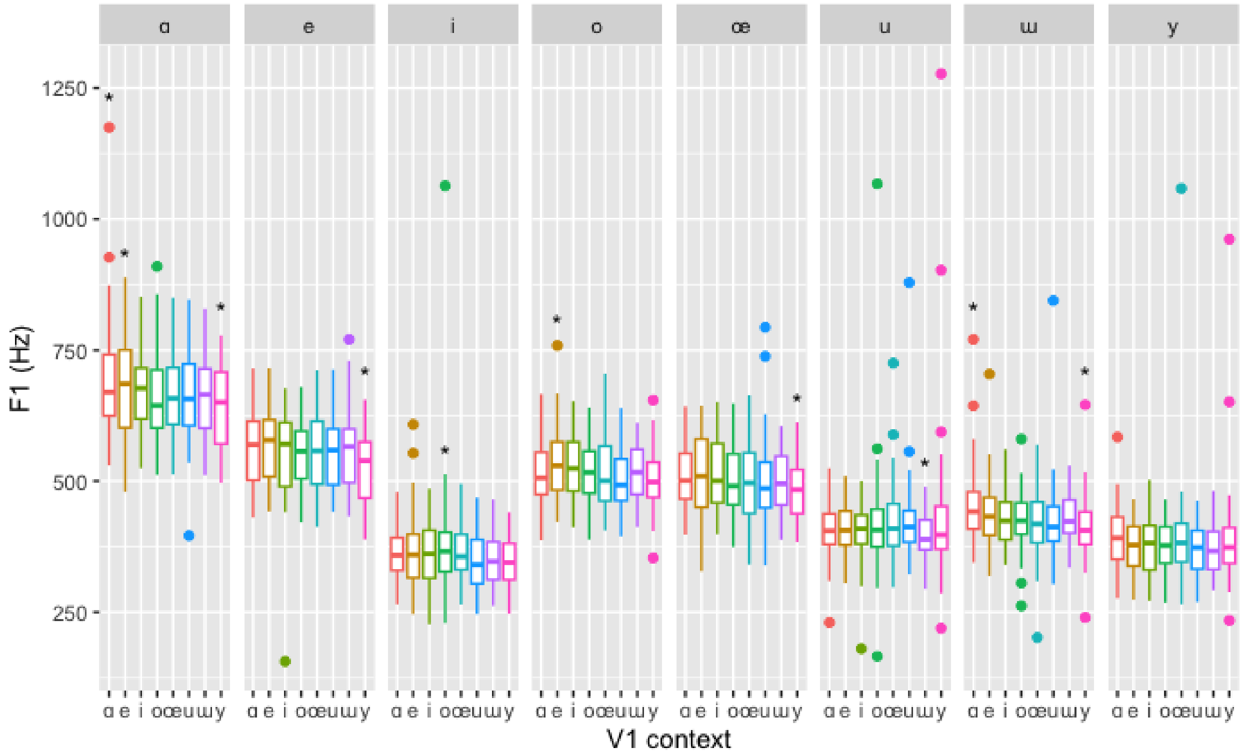

Figure 2.15. Vowel F1 at vowel onset for V_2 vowels. Asterisks indicate significant differences from the V_2 vowel mean F1.

In the F2 model, there was a significant interaction of V_1 vowel category and V_2 vowel category ($F(63, 4964) = 394.67$, $p < 0.0001$), indicating that vowels exhibited shifts in F2 that varied by target vowel category and context vowel category. Planned comparisons revealed several significant anticipatory coarticulation effects in 10 of the pairwise comparisons out of 64 (Table 2.14). These significant shifts in vowel F2 at V_2 vowel onset are in line with the predictions of carryover coarticulation whereby lower F2 in V_1 caused V_2 to back and higher F2 in V_1 caused V_2 to front, with the exception of V_2 /e/ fronting in the context of V_1 /ɑ/, and V_2 /u/ backing in the context of V_1 /ɑ/ and V_1 /e/ (Figure 2.16).

Table 2.14. Significant pairwise comparisons in the carryover coarticulation in vowel F2 model.

V_2 Vowel (target)	V_1 Vowel (context)	Estimate (ΔF2)	SE	z	p
/ɑ/	/o/	-60.76	18.7	-3.24	0.01*
/e/	/ɑ/	55.76	18.7	2.98	0.02*
/ɯ/	/o/	-89.22	18.7	-4.76	< 0.0001***
	/y/	80.64	18.7	4.3	0.0001***
/œ/	/e/	72.43	18.7	3.87	0.001**
/u/	/ɑ/	-54.63	18.7	-2.91	0.03*
	/e/	-51.65	18.7	-2.76	0.047*
	/œ/	51.76	18.7	2.75	0.048*
	/y/	103.67	18.7	5.53	< 0.0001***
/y/	/œ/	-57.35	18.7	-3.06	0.02*

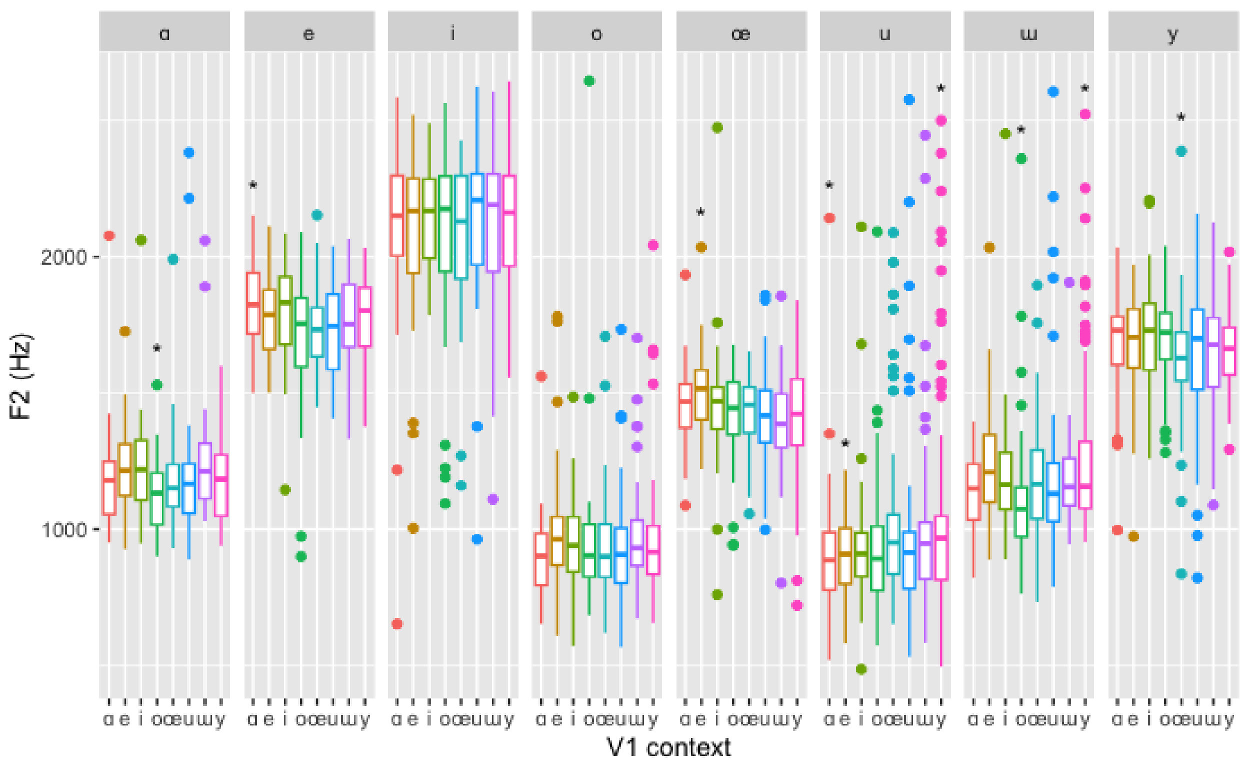

Figure 2.16. Vowel F2 at vowel onset for V_2 vowels. Asterisks indicate significant differences from the V_2 vowel mean F2.

Overall, these results support carryover coarticulation effects in Turkish vowels (Figure 2.17), although significant formant shifts are observed only for a limited number of cases that do not make a general pattern. Where there was a significant effect, in general, high V_1 vowels caused V_2 vowels to raise, non-high V_1 vowels caused V_2 vowels to lower, front V_1 vowels caused V_2 vowels to front, and back V_1 vowels caused V_2 vowels to back. However, these significant formant shifts are observed only for a limited number of cases and do not generalize over vowel feature classes. In contrast to anticipatory coarticulation, both peripheral and nonperipheral vowels participated in carryover coarticulation as targets and triggers.

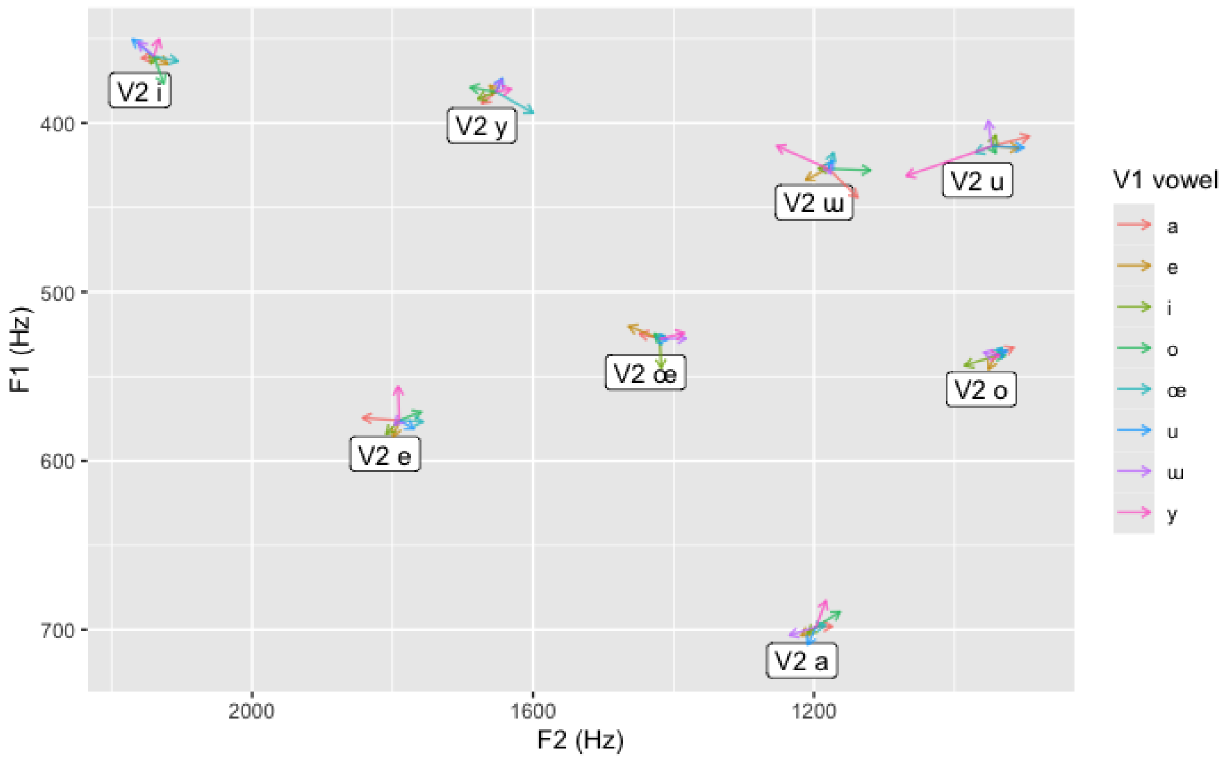

Figure 2.17. Carryover coarticulation effects on V_2 vowels by V_1 vowel context in disyllabic nonwords.

The findings here suggest that carryover coarticulation effects are stronger and more widespread compared to anticipatory coarticulation effects. These results diverge from previous studies suggesting that anticipatory coarticulation is stronger than carryover coarticulation in Turkish (Beddor & Yavuz, 1995; Inkelas et al., 2001), which examined /ɑ e i/ only as targets and triggers. I argue that the weaker and smaller anticipatory coarticulation effects I find are in line with the vowel position findings in this study: vowels in word-initial syllables are significantly longer and more hyperarticulated, suggesting that they are less likely to be targeted by anticipatory coarticulation and more likely to be triggers of carryover coarticulation. Yet, carryover and anticipatory coarticulation effects alike were rare, suggesting a general tendency for vowels to not coarticulate.

2.4. Interim discussion

The inspection of the cardinal Turkish vowels in monosyllabic nonwords revealed that Turkish vowels are phonetically grouped into 3 height classes and are continuously distributed along the F2 dimension. These results suggest that the phonological vowel features and harmony classes do not directly map onto the Turkish vowel space. In particular, the phonetic vowel space does not have clear distinctions corresponding to vowel backness and rounding as described phonologically. I argue that cues to rounding such as F3 or coarticulatory lip rounding in tautosyllabic consonants (Boyce, 1990) might be perceptually salient. Comparisons of Turkish vowels across syllable positions suggested that vowels in V_1 position might be longer and more hyperarticulated, suggesting that vowels in word-initial syllables might be phonologically more prominent (Barnes, 2002). Hyperarticulation due to vowel position was more widespread in the F2 dimension, presumably to facilitate contrast enhancement for the backness and rounding features in particular for the central or nonperipheral vowels.

I found limited evidence for vowel-to-vowel coarticulation, although carryover coarticulation effects were stronger and more widespread than anticipatory coarticulation effects. That vowel-to-vowel coarticulation patterns were specific to individual vowel pairs suggests that Turkish vowels participate in vowel-to-vowel coarticulation as individual vowels rather than as harmony classes. In addition, I found that nonperipheral vowels /ɯ œ y/ exhibit less variation in vowel formants caused by hyperarticulation and vowel-to-vowel coarticulation compared to peripheral vowels /ɑ i o u/. I argue that nonperipheral vowels might be perceptually disadvantaged compared to peripheral vowels with more extreme F1-F2 values, and hence exhibit less phonetic variation to facilitate contrast preservation. In the following sections, I examine allophonic variation in Turkish vowels and ask whether they exhibit patterns that are aligned with harmony classes.

2.5. Turkish Vowel Allophones

There is no agreement on what the allophones of the 8 canonical Turkish vowels are in the literature. The most comprehensive lists of proposed Turkish vowel allophones are found in descriptive grammars of the Turkish language which do not reference production data or phonetic measurements (see Yılmaz Davutoğlu, 2010 for a review). Among this body of work, Özsoy's (2004) work is the most extensive, which posits 17 vowel allophones in total and specifies the phonological contexts, proposes phonological derivation rules for the surface forms, and provides numerous examples for each allophone. For the 8 Turkish vowels, Özsoy identifies 8 nasal allophones, 8 lower allophones, and additionally [ɛ] as an allophone of /e/. According to Özsoy, when a vowel is followed by a tautosyllabic nasal consonant, the vowel surfaces as its nasal allophone. When a vowel (1) occurs in a word-final open syllable or (2) is followed by tautosyllabic sonorant consonant, the vowel surfaces as its lower allophone. These two contexts result in two separate allophones for the vowel /e/: in (1), /e/ surfaces as a lowered /e/, whereas in (2), /e/ surfaces as lowered and backed [ɛ]. Lastly, Özsoy posits that vowel length is phonemic in Turkish and that short and long vowels have the same phonetic makeup except for their durations.

A comparable descriptive work with no phonetic measurements is by Göksel and Kerslake (2005), who identify 8 vowel allophones, specifically fronted allophones for 3 of the back vowels /ɑ o u/ in the context of palatalized consonants [c ɟ l], 4 lower or lax allophones in word-final open syllables including [ɛ] and lower allophones of /i u y/, and [æ] as an allophone of /e/ followed by tautosyllabic sonorant consonants. Similarly, Erguvanlı-Taylan (2015) identifies 4 vowel allophones, specifically the same 3 fronted vowel allophones and [ɛ] when /e/ is followed by a sonorant coda. Note that whereas Özsoy (2004) and Erguvanlı-Taylan agree on the /e/ allophone in sonorant coda contexts being an [ɛ], Göksel and Kerslake posit that [ɛ] is the allophone in word-final open syllables and [æ] is the allophone in sonorant coda contexts. Hence, aside from the disagreement in the set of Turkish vowel allophones, researchers also disagree at times in the phonetic identity or transcription of the same allophone in their descriptions of the Turkish vowel system. Lastly, unlike Özsoy, neither Göksel and Kerslake nor Erguvanlı-Taylan comment on the phonetic makeup of long vowels, although they discuss how the vowel length contrast is best analyzed phonologically in Turkish.

Finally, in an attempt to describe the entire Turkish sound system, Ergenç and Uzun (2020) posit 7 allophones consisting of allophones based on jaw aperture for 6 Turkish vowels (including [ɛ] as an allophone of /e/ but excluding lower allophones of /ɑ/ and /ɯ/) and one palatalized allophone, [a], for the vowel /ɑ/. The authors also specify that the 8 canonical vowels of Turkish are short, and that short vowels are open whereas long vowels or vowels that are diphthongized due to a following <ğ> or /j/ (see below) surface as the close allophone of the vowel. One exception to this description is /e/: both close /e/ and its open allophone [ɛ] are claimed to be short and long, with the allophonic variation being conditioned by vowel position whereby /e/ vowels in noninitial position or open syllables surface as [ɛ]. Similarly, both /ɑ/ and its palatalized allophone [a] can be both short and long. As such, Ergenç and Uzun's work disagrees with Özsoy (2004) in that long vowels exhibit allophonic variation. Although the authors obtain phonetic measurements from an unspecified number of native Turkish speaker productions, they only provide visualizations of certain measures such as vowel formant tracks and spectrograms for a small number of vowel allophone productions without reporting measured values. Hence, Ergenç and Uzun's study is grouped with other work that is descriptive of the Turkish sound system.

Together, the selected references reviewed here suggest that nasalization (although see Kornfilt, 1997, who argues that vowel nasalization is not perceptible unless the nasal coda is deleted and the vowel is compensatorily nasalized), vowel lowering in word-final open syllables and in syllables with sonorant codas, vowel fronting in the context of palatalized consonants, vowel raising in long vowels, and [ɛ] constitute the potential allophones of the 8 Turkish vowels and are worth examination. All proposed allophones are conditioned by the context following the vowel, with the exception of vowel fronting, which is proposed to occur in both preceding and following palatalized consonant contexts. A small number of Turkish vowel production studies in the literature are concerned with identifying Turkish vowel allophones and confirmatory phonetic analyses of the proposed Turkish vowel allophones. However, the vowel allophones examined in these studies do not exhaust the vowel allophones that are proposed in the work reviewed above.

An extensive set of proposed Turkish vowel allophones is examined with a relatively large production dataset in Yılmaz Davutoğlu (2010). The author reviews the literature to identify an exhaustive list of proposed Turkish vowel and consonant allophones to create a word list including all proposed allophones, controlling for word-position, stress, and vowel length. The resulting 875-word list was produced by 8 Turkish speakers, yielding a total of 16,288 vowel tokens. Yılmaz Davutoğlu's approach to Turkish vowel allophones is bottom-up as it identifies the individual contexts that are predicted to yield differences in F1 or F2 values for vowels and groups the contexts according to their predicted effects for statistical analysis. The author identifies 13 statistically significant vowel allophones and 10 long vowel variants in this manner via t-tests and ANOVAs and reports mean and standard deviation F1-F2 values for the identified allophones along with a description of the contexts in which these allophones are observed. However, it is not clear whether some of the identified allophones are better described as coarticulation due to various neighboring consonants that are not necessarily defined by shared phonological features. Moreover, neither the exhaustive list of hypothesized allophones nor the list of words is reported, which obscures some of the details of the statistical analyses. Despite its limitations, Yılmaz Davutoğlu's work is nonetheless impressive in its scope.

Comparably, Börtlü (2020) reviews the descriptive literature on Turkish vowel allophones and selects 4 allophones, specifically fronted allophones of /ɑ o u/ and [ɛ], for acoustic analysis to describe the Turkish vowel space, although it is not discussed why only these 4 allophones previously proposed in the literature were selected. A list of 51 words where the 12 target vowel sounds occurred word-initially, word-medially, and word-finally were produced by 10 native Turkish speakers, yielding 960 vowel tokens in total. The vowels occurred in the contexts of plosives, fricatives, and sonorants, although the author does not discuss consonantal effects on vowel production. By-speaker mean F1 and F2 values at vowel midpoint are reported for the 12 vowel sounds, and the raw and normalized vowel formants are visualized as vowel clouds on F1-F2 space. Although no statistical analyses are performed, the author discusses in detail the environments in which the 4 allophones investigated in their study surface to identify phonological rules predicting the complementary distributions of the Turkish vowels and their allophones. The author identifies tautosyllabic palatalized consonants /g k l/ as the environment that licenses vowel fronting and argues that despite not being attested in the Turkish lexicon, the same environment is predicted to cause vowel fronting for /ɯ/ is well. The author identifies a set of rules for /e/ surfacing as [ɛ], which account for tautosyllabic sonorant consonant context, word position, syllable number, and the syllable-initial consonant of the following syllable. Börtlü's study is arguably richer in its phonological analysis of the exceptional environments in which the vowels do or do not surface as their respective allophones than it is in the phonetic analysis that accompanies these observations.

Another study that is concerned with back vowel fronting is by Canalis and Dikmen (2020). In this production study, native Turkish speakers produced near minimal pairs with palatalized and non-palatalized allophones of consonants preceding and following back vowels /ɑ o/. The authors compared vowel F2 in stem-final syllables across near minimal pairs and found that vowels had significantly higher F2 in the context of palatalized allophones of consonants. Moreover, vowels exhibited F2 movement throughout the vowel such that if the palatalized allophone was preceding the vowel, F2 was higher in vowel onset than in vowel midpoint whereas if the palatalized allophone was following the vowel, F2 was higher in vowel offset than in vowel midpoint. Thus, the authors argue that back vowel fronting in Turkish is a consonant-to-vowel coarticulatory phenomenon.

An exceptional study by Dadan et al. (2024) aims to validate proposed Turkish vowel allophones and focuses on the acoustic analysis of the allophones of a single vowel, /e/. The authors collected a total of 4200 /e/ tokens in various environments licensing allophones from 9 native Turkish speakers and analyzed vowel F1 via linear mixed effects modeling to test the two allophones of /e/ ([æ] and [ɛ]) that Göksel and Kerslake (2005) hypothesize. Their results confirm that /e/ is lowered in the context of sonorant codas and lowered to a lesser extent in word-final open syllables, which the authors phonetically classify as [æ] and [ɛ], respectively. Dadan et al. additionally show that /e/ is realized as [æ] in the word *pekmez* and the negative aorist *-mez* but not in other *mez* syllables or syllables with /z/ codas (cf. Gopal & Nichols, 2022).

Overall, the work reviewed above accumulates a considerable set of proposed Turkish vowel allophones, which have not been exhaustively confirmed via phonetic analyses. In the following sections, I will present data comparing vowel phonetics in Turkish words to examine whether vowel lengthening, vowel fronting, nasalization, sonorant codas, and word-final open syllable vowel lowering yield vowel allophones that are phonetically distinct from the phonemic, canonical vowels. I ask whether vowels within a harmony class behave similarly in allophonic variation.

2.5.1. Vowel length

The picture regarding the understanding of vowel length in Turkish is similar to that of Turkish vowel allophones—descriptive grammars are in disagreement and the few phonetic analyses that are present in the literature are insufficient (see Ünal-Logacev et al., 2019, for a review). Most descriptive grammars of Turkish identify two separate long vowel phenomena (Ergenç & Uzun, 2020; Erguvanlı-Taylan, 2015; Göksel & Kerslake, 2005; Kornfilt, 1997; Özsoy, 2004), the first type observed in mainly Arabic and Persian borrowings in which the original word has a long vowel /aː eː iː uː/ and hence the vowel is considered to be phonemically long (Kornfilt, 1997) and the other type observed in native Turkish words via a phonological process of compensatory lengthening as will be described in more detail. All 8 Turkish vowels can be long and make minimal pairs with respect to vowel length in native Turkish words, as exampled in (9):

(9)

Short vowel	Long vowel
arı [ɑ.ɾɯ] – ‘bee’	ağrı [ɑː.ɾɯ] – ‘pain’
sıla [sɯ.ɫɑ] – ‘homeland’	sığla [sɯː.ɫɑ] – ‘sweetgum’
olan [o.ɫɑn] – ‘happening’	oğlan [oː.ɫɑn] – ‘boy’
bura [bu.ɾɑ] – ‘here’	Buğra [buː.ɾɑ] or [buː.ɾɑː] – (a male name)
ye [ye] – ‘eat (imperative)’	yeğ [yeː] – ‘preferable’
emek [e.mek] – ‘labor’	eğmek [eː.mek] – ‘to bend’
ine [i.ne] – ‘descend (imperative)’	iğne [iː.ne] – ‘needle’
ölen [œ.lɛn] – ‘dying’	öğlen [œː.lɛn] – ‘noon’
dümen [dy.mɛn] – ‘rudder’	düğmen [dyː.mɛn] – ‘your button’

Notice that in all long vowel words, the orthographic representation of the word has the Turkish grapheme <ğ> ‘soft g’ following the vowel. In the phonetic transcriptions, there is no consonantal sound corresponding to the syllable-final ‘soft g’ in orthography and yet the minimal pairs differ in that the vowels preceding the ‘soft g’ are long. According to most descriptive grammars of Turkish (Erguvanlı-Taylan, 2015; Göksel & Kerslake, 2005; Kornfilt, 1997; Özsoy, 2004), ‘soft g’ is a remnant of voiced velar fricative /ɣ/ that is no longer part of the sound system of Turkish. When in coda position ‘soft g’ is phonetically zero due to deletion of the archaic voiced velar fricative and lengthens the vowel it follows as a result of compensatory lengthening. However, when vowel lengthening ‘soft g’ is word-final, it is treated like a consonant in suffixation (e.g., the accusative suffix allomorph /-ɯ/ surfaces with an intervening consonant if the root has a word-final vowel: <dal> [dɑɫ] and <dağ> [dɑː] become [dɑɫ-ɯ] ‘branch-ACC’ and [dɑ-ɯ] ‘mountain-ACC’, respectively, whereas <ada> [ɑdɑ] becomes [ɑdɑ-jɯ] ‘island-ACC’ and an Arabic borrowing with a phonemically long vowel in word-final position such as <eda> [edaː]

becomes [edaː-jı] 'manner- ACC')[1]. Due to its consonant-like behavior in word-final position, some have argued that 'soft g' is phonemic (Kabak, 2007) or is an abstract consonant (Erguvanlı-Taylan, 2015; Kornfilt, 1997; Lees, 1961; Ünal-Logacev et al., 2019).

There is disagreement in the descriptive literature whether 'soft g' is acoustically realized as a velar, semi-vowel or glide intervocalically or in the context of front vowels or rounded vowels (see Ünal-Logacev et al., 2019 for a review, and Taşdemir et al., 2025 for a preliminary phonetic analysis of 'soft g' in various contexts). In contrast, whether the phonetically zero abstract coda consonant nonetheless effects the phonetic makeup of the preceding vowel other than in vowel length has not received much discussion. Whereas most descriptive grammars do not discuss the phonetic realization of lengthened vowels, Özsoy (2004) claims that the lengthened vowels are phonetically identical with their short counterparts except for their duration. The question of the phonetic realization of the long vowels has not been a subject of phonetic analysis either.

There is a small number of production studies involving 'soft g' in Turkish, however these studies are focused on the 'soft g' rather than the vowels in its environment. Coşkun (2000) anecdotally reports observing identical F1 and F2 values when comparing vowel and vowel-'soft g' sequence minimal pairs, agreeing with Özsoy (2004) in effect. In contrast, Kılıç and Erdem (2013) observe that F1 lowers throughout the production of the vowel in the word <dağ> [dɑː], and more generally claim that the 'soft g' is associated with a velar gesture, which lowers the

[1] Although Kabak, 2007 posits that <bağ> [bɑː] becomes [bɑ-ɯ] 'vineyard-ACC' and not [bɑː-ɯ] as 'soft g' is resyllabified to onset position due to suffixation, he perhaps mistakenly transcribes <dağı> 'mountain-ACC' as [dɑːɯ]. I prefer [dɑ-ɯ] and [bɑ-ɯ] to differentiate between 'soft g' in coda position and intervocalic position, however without a commitment to the acoustic realization of intervocalic 'soft g' or the effect it has on the contextual vowels, which has not been investigated sufficiently. Taşdemir et al.'s (2025) preliminary phonetic analysis suggests that, in contrast to the coda 'soft g' which lengthens the preceding vowel, intervocalic 'soft g' might lengthen the following vowel. Also note that Kabak transcribes an Arabic borrowing <kirayı> 'rent-ACC' as [cira-jɯ], whereas I prefer [ciraː-jɯ] as in [edaː-jɯ], as the root-final vowels are phonemically long in these words.

neighboring vowels' F1, and raises the neighboring vowels' F2 when the vowel is front and lowers it when the vowel is back. However, both of these studies are lacking a substantial phonetic analysis.

Ünal-Logacev et al.'s (2019) study on 'soft g' in Turkish is the sole study statistically analyzing phonetic measurements of productions. In their phonetic analysis, the authors do not find any phonetic evidence for a velar gesture for 'soft g' in any position. Ünal-Logacev et al. compare minimal pairs with and without a following coda 'soft g' for /ɑ/ and /e/ only, and report higher F1, 'more peripheral positions' and 'a larger vowel space' in addition to increased vowel length for the vowel preceding the 'soft g' in coda position. The authors also phonetically analyze identical vowel sequences with an intervocalic 'soft g' in words such as <kuğu> 'swan' and similarly find no acoustic evidence of a velar gesture. Hence, they suggest that intervocalic 'soft g' is phonetically zero and the identical vowel sequence is produced as one long vowel. The authors find that F1 keeps increasing throughout the production of these long vowels, which contrasts with Kılıç and Erdem's (2013) finding that 'soft g' lowers the F1 of the neighboring vowel. Ünal-Logacev et al. attribute the increase in F1 in their data to longer vowel production allowing tongue movement associated with more peripheral productions.

Considering there are only a few production studies on 'soft g' in Turkish, which do not present converging findings on how long vowels in the context of 'soft g' are realized phonetically, I set out with a broad hypothesis that these long vowels are phonetically different from their short counterparts in minimal pairs in F1 and F2.

2.5.1.1. Stimuli

Among the list of Turkish words the participants produced in carrier sentences, 14 had long vowels in the first syllable due to a following ‘soft g’ which made minimal pairs or near minimal pairs (Table 2.15). The 26 speakers produced 2 tokens of each word of the 14 (near) minimal pairs, yielding a total of 728 long vowel tokens that are being compared to a total of 728 short vowel tokens.

Table 2.15. Minimal or near minimal pairs by vowel length in the production study.

Target vowel	Vowel length	Target word	Phonetic transcription	Gloss
ɑ	Short	arı	ɑ.ɾɯ	bee
	Long	ağrı	ɑː.ɾɯ	pain
ɑ	Short	kanı	kɑ.nɯ	opinion
	Long	kağnı	kɑː.nɯ	ox cart
e	Short	emek	e.mek	labor
	Long	eğmek	eː.mek	to bend
e	Short	demek	de.mek	to say
	Long	değmek	deː.mek	to touch
e	Short	ye	ye	eat
	Long	yeğ	yeː	preferable
ɯ	Short	ıtır	ɯ.tɯɾ	geranium
	Long	Iğdır	ɯː.dɯɾ	(a city in Turkey)
i	Short	inek	i.nec	cow
	Long	iğne	iː.ne	needle
o	Short	olan	o.ɫɑn	happening
	Long	oğlan	oː.ɫɑn	boy
o	Short	doru	do.ɾu	chestnut, bay
	Long	doğru	doː.ɾu	true
œ	Short	ölen	œ.lɛn	dying
	Long	öğlen	œː.lɛn	noon
œ	Short	börü	bœ.ɾy	wolf
	Long	böğrü	bœː.ɾy	belly-ACC
u	Short	bura	bu.ɾɑ	here
	Long	Buğra	buː.ɾɑ	(a boy’s name)
u	Short	bu	bu	this
	Long	buğu	buː (or buu)	vapor
y	Short	dümen	dy.mɛn	rudder
	Long	düğme	dyː.me	button

2.5.1.2. Data analysis

The phonetic measures of interest were vowel duration and F1-F2 at vowel midpoint. Each of these measures were modeled as the dependent variable in separate linear mixed effects models with vowel category, vowel length, speaker gender, and the interaction of vowel category and vowel length as predictors. There were random intercepts for speaker and minimal pair, with vowel length as a random slope for minimal pair (10). Multi-level fixed effects were evaluated via a type III ANOVA using Satterthwaite's Method. Vowel category was sum coded. Vowel length had two levels (short and long) and was sum coded (short: -1, long: 1) for ease of interpretation. Planned comparisons between levels of a variable were performed using the `emmeans` package (Lenth & Piaskowski, 2025) with Bonferroni adjustment for multiple comparisons.

(10) vowel duration or formant ~ vowel category
+ vowel length
+ speaker gender
+ vowel category : vowel length
+ (1 | speaker)
+ (1 + vowel length | minimal pair)

2.5.1.3. Results

In the duration model, there was a significant main effect of vowel category ($F(7, 13.87) = 4.51$, $p = 0.008$), indicating that vowel duration varied by vowel category. There was a significant main effect of vowel length ($\beta = 46.16$, $SE = 1.37$, $t = 33.80$, $p < 0.0001$), confirming that vowels followed by 'soft g' were longer than vowels in open syllables. There was no significant main effect of gender (*p* = 0.84, Table 2.16). There was a significant interaction of vowel category and vowel ($F(7, 14.00) = 4.39$, $p = 0.009$), indicating that the difference between short and long vowel durations varied across vowel categories. Planned comparisons revealed that for all vowel categories, the difference between short and long vowel durations were significant (all *p*s ≤ 0.0001; Figure 2.18). Hence, this model confirmed that 'soft g' in coda position significantly lengthens preceding vowels and that all 8 Turkish vowels undergo vowel lengthening due to 'soft g' in coda position. In the following models, I examine whether vowel lengthening influences vowel formants. In other words, I examine whether long vowels' formant structure differs from that of short vowels.

Table 2.16. Duration model results.

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	129.89	4.96	28.95	26.18	< 0.0001***
/ɑ/	13.33	9.19	13.87	16438	0.17
/e/	-3.06	7.83	13.87	-0.39	0.70
/i/	5.88	12.39	13.87	0.47	0.64
/o/	27.61	9.19	13.87	45660	0.01**
/œ/	21.41	9.19	13.87	12086	0.04*
/u/	6.17	9.19	13.87	0.67	0.51
/ɯ/	-18.16	12.39	13.87	-1.47	0.16
Vowel length	46.16	1.37	14.00	33.80	< 0.0001***
Speaker gender	-0.63	3.14	25.71	-0.20	0.84
/ɑ/: Vowel Length	7.06	3.20	14.00	43862	0.04*
/e/: Vowel length	-0.42	2.73	14.00	-0.16	0.88
/i/: Vowel length	5.95	4.32	14.00	13881	0.19
/o/: Vowel length	-8.47	3.20	14.00	-2.64	0.02*
/œ/: Vowel length	-2.84	3.20	14.00	-0.89	0.39
/u/: Vowel length	-0.10	3.20	14.00	-0.03	0.97
/ɯ/: Vowel length	13.45	4.32	14.00	45994	0.007**

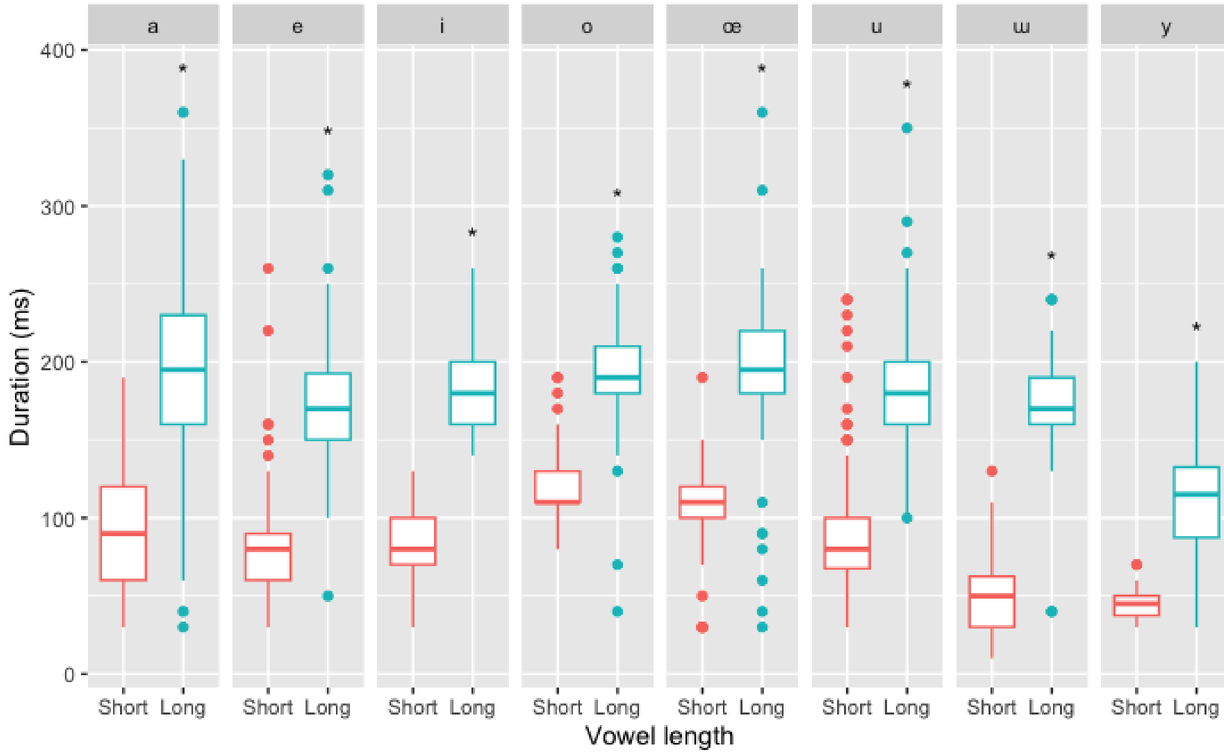

Figure 2.18. Vowel duration in short and long vowels in (near) minimal pair words. Asterisks indicate significant differences in vowel length across short and long vowels.

In the F1 model (Table 2.17), the dependent variable was F1 at vowel midpoint. There was a significant main effect of vowel category ($F(7, 27.68) = 60.57$, $p = 4.079\text{e-}15$), indicating that controlling for vowel length, vowel category was a significant predictor of vowel F1. Vowel length did not have a significant main effect on vowel F1 across the vowel categories ($\beta = -8.42$, $SE = 4.21$, $t = -2.00$, $p = 0.07$), however, there was a significant interaction of vowel category and vowel length ($F(7, 14.00) = 9.63$, $p = 0.0002$), indicating that vowel length had an effect on vowel F1 that varied by vowel category (Figure 2.19). Planned comparisons (Table 2.18) confirmed the model results that long /ɑ/ vowels had significantly higher F1 than short /ɑ/ vowels whereas long /e/ and /ɯ/ vowels had significantly lower F1 than their short counterparts. Lastly, there was a significant main effect of gender ($\beta = 55.22$, $t = 8.98$, $p = 1.97\text{e-}09$), indicating that female speakers' vowels had higher F1 values compared to male speakers controlling for vowel length.

Table 2.17. F1 model results.

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	460.43	8.38	32.24	54.96	< 0.0001***
/ɑ/	211.13	14.21	13.88	14.86	< 0.0001***
/e/	75.75	12.12	13.88	6.25	< 0.0001***
/i/	-122.16	19.16	13.88	-6.38	< 0.0001***
/o/	14.08	14.21	13.88	0.99	0.34
/œ/	3.75	14.21	13.88	0.26	0.80
/u/	-80.01	14.21	13.88	-5.63	< 0.0001***
/ɯ/	2.51	19.16	13.88	0.13	0.90
Vowel length	-8.42	4.21	14.00	-2.00	0.07
Speaker gender	55.22	6.14	25.92	9.00	< 0.0001***
/ɑ/: Vowel Length	63.45	9.88	14.00	6.42	< 0.0001***
/e/: Vowel length	-24.88	8.42	14.00	-2.95	0.01*
/i/: Vowel length	-4.21	13.32	14.00	-0.32	0.76
/o/: Vowel length	13.54	9.88	14.00	1.37	0.19
/œ/: Vowel length	17.63	9.88	14.00	1.78	0.10
/u/: Vowel length	-4.78	9.88	14.00	-0.48	0.64
/ɯ/: Vowel length	-59.19	13.32	14.00	-4.44	< 0.0001***

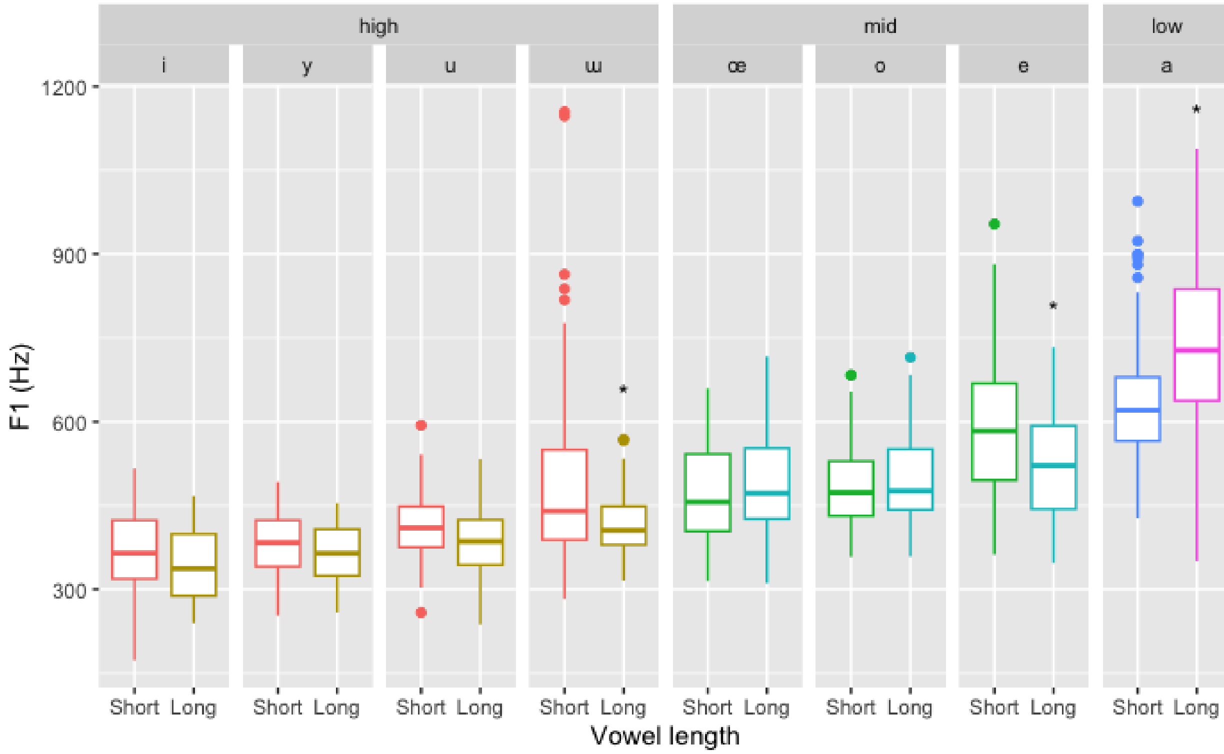

Figure 2.19. Vowel F1 in short and long vowels in (near) minimal pair words. Asterisks indicate significant differences in vowel F1 across short and long vowels.

Table 2.18. Significant pairwise comparisons in the vowel F1 by vowel length model.

Vowel	Estimate (ΔF1)	SE	t	p
/ɑ/	110.1	31.5	3.49	0.001**
/e/	-66.6	25.7	-2.98	0.01*
/ɯ/	-135.2	44.6	-3.03	0.005**

In the F2 model (Table 2.19), the dependent variable was F2 at vowel midpoint. There was a significant main effect of vowel category ($F(7, 13.65) = 858.47$, $p < 0.0001$), indicating that controlling for vowel length, vowel category was a significant predictor of vowel F2. Vowel length did not have a significant main effect on vowel F2 across the vowel categories ($\beta = -15.23$, $SE = 8.55$, $t = -1.78$, $p = 0.1$), however, there was a significant interaction of vowel category and vowel length ($F(7, 14.00) = 14.44$, $p < 0.0001$) indicating that vowel length had an effect on vowel F2 that varied by vowel category (Figure 2.20). Planned comparisons (Table 2.20) confirmed model results that long /ɑ/ and /u/ vowels had significantly lower F2 than short /ɑ/ and /u/ vowels and long /e/ vowels had significantly higher F2 than short /e/ vowels. In addition, long /ɯ/ vowels had marginally significantly lower F2 than short /ɯ/ vowels and long /y/ vowels had marginally significantly higher F2 than short /y/ vowels. Lastly, there was a significant main effect of gender, indicating that female speakers' vowels had higher F2 values compared to male speakers controlling for vowel length ($\beta = 91.97$, $t = 8.40$, $p < 0.0001$).

Table 2.19. F2 model results.

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	1508.19	11.66	26.40	129.35	< 0.0001***
/ɑ/	-264.69	15.48	13.56	-17.10	< 0.0001***
/e/	478.03	13.20	13.56	36.22	< 0.0001***
/i/	913.90	20.87	13.56	43.79	< 0.0001***
/o/	-606.63	15.48	13.56	-39.20	< 0.0001***
/œ/	-109.17	15.48	13.56	-7.05	< 0.0001***
/u/	-592.14	15.48	13.56	-38.26	< 0.0001***
/ɯ/	-31.12	20.87	13.56	-1.49	0.16
Vowel length	-15.23	8.55	14.00	-1.78	0.09
Speaker gender	91.97	10.94	25.99	8.40	< 0.0001***
/ɑ/: Vowel Length	-88.67	20.04	14.00	-4.42	0.0006**
/e/: Vowel length	106.78	17.09	14.00	6.25	< 0.0001***
/i/: Vowel length	85.33	27.02	14.00	3.16	0.007**
/o/: Vowel length	-38.34	20.04	14.00	-1.91	0.08
/œ/: Vowel length	-21.95	20.04	14.00	-1.10	0.29
/u/: Vowel length	-74.62	20.04	14.00	-3.72	0.002**
/ɯ/: Vowel length	-73.25	27.02	14.00	-2.71	< 0.02*

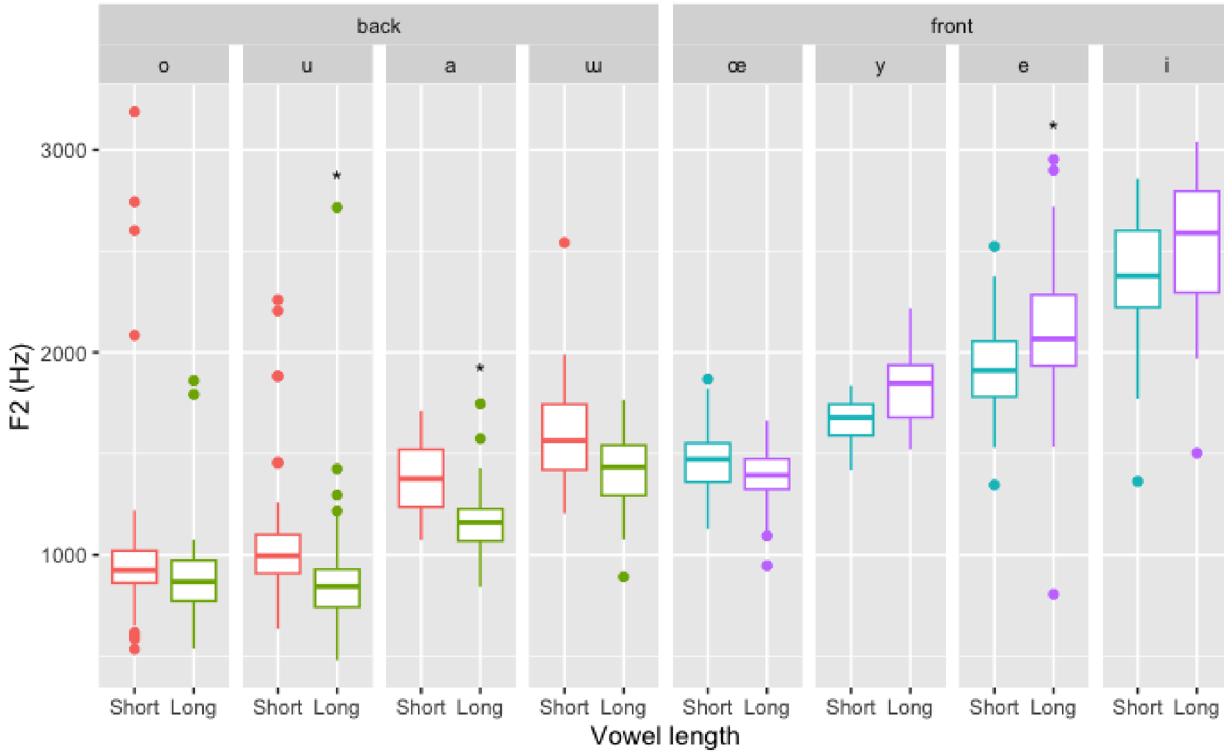

Figure 2.20. Vowel F2 in short and long vowels in (near) minimal pair words. Asterisks indicate significant differences in vowel F2 across short and long vowels.

Table 2.20. Significant and marginally significant pairwise comparisons in the vowel F2 by vowel length model.

Vowel	Estimate (ΔF2)	SE	t	p
/ɑ/	-207.8	64.0	-3.25	0.003*
/e/	183.1	52.2	3.51	0.001*
/ɯ/	-177.0	90.4	-1.96	0.059
/u/	-179.7	64.0	-2.81	0.008*
/y/	179.0	90.4	1.98	0.056

2.5.1.4. Discussion

The investigation of vowels in (near) minimal pairs with and without coda ‘soft g’ confirmed the vowel lengthening status of ‘soft g’ in coda position. Comparisons of short and long vowel F1-F2 values revealed that most lengthened vowels also undergo changes in their formant structures as observed in an interaction of vowel category and vowel length in the models (Figure 2.21). Overall, these results do not agree with Özsoy's (2004) descriptive grammar of Turkish which claims that long and short vowels are phonetically identical except for their durations.

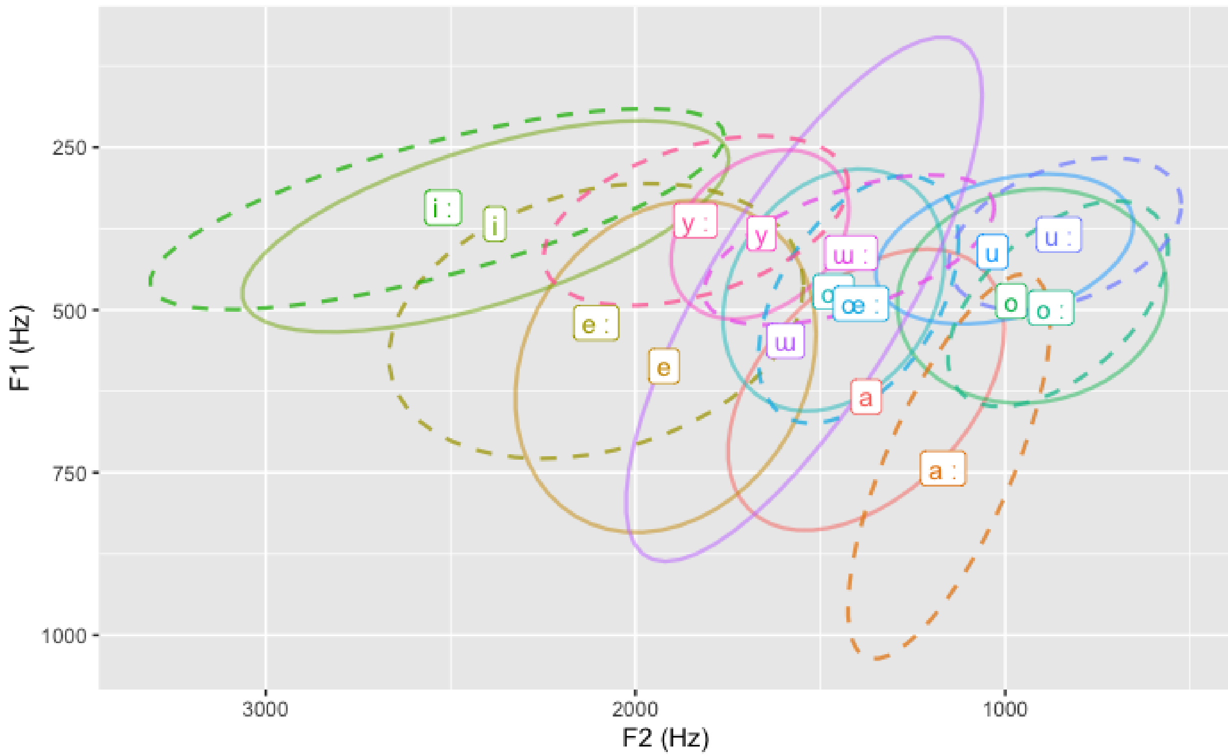

Figure 2.21. Vowel ellipses for short vowels (solid ellipses) and long vowels (dashed ellipses) in (near) minimal word pairs

Pairwise comparisons summarized in Table 2.21 revealed both vowel raising and lowering and vowel backing and fronting in long vowels. The bidirectional changes in F1 observed here contrast with Kılıç and Erdem (2013) who claim that long vowels have lower F1 and with Ünal-Logacev et al., (2019) who claim that long vowels have higher F1 compared to short vowels. In contrast, the bidirectional changes in F2 observed here are in line with Kılıç and Erdem's (2013) claim that long back vowels have lower F2 than short back vowels whereas long front vowels have higher F2 than short front vowels. Moreover, pairwise comparisons suggested that when there are formant changes in long vowels, the changes are generally in line with hyperarticulation and contrast enhancement, such as back vowels backing when long. As such, the findings here parallel the findings in the previous sections showing longer and more hyperarticulated vowel production in V_1 position relative to V_2 position. In addition, I find that long /e/ vowels raised when long, paralleling /e/ raising in V_1 position. As argued above, as /e/ is phonologically non-high and phonetically mid, /e/ raising is not predicted under hyperarticulation or contrast enhancement, yet can be attributed to contrast preservation between /e/ and its lower allophone [ɛ].

Table 2.21. Summary of the vowel length effects on vowel F1 and F2 in the pairwise comparisons.

Vowel	Height	F1 in long vowels	Backness	F2 in long vowels
/ɑ/	Non-high (low)	Lower	Back	Backer
/e/	Non-high (mid)	Higher	Front	Fronter
/ɯ/	High		Back	Marginally backer
/i/	High	Higher	Front	
/o/	Non-high (mid)		Back	
/œ/	Non-high (mid)		Front	
/u/	High		Back	Backer
/y/	High		Front	Marginally fronter

It should be noted that the significant changes in F1 or F2 in long vowels here do not apply to vowels as a height or backness class entirely. Among high vowels only /i/ exhibited raising and among non-high vowels only /ɑ/ exhibited lowering. Among back vowels /ɑ/ and /u/ exhibited significant backing and among front vowels only /e/ exhibited significant fronting. Similar to the findings suggesting more hyperarticulated vowel productions in V_1 position compared to V_2 position, the findings from vowel length comparisons suggest vowels are more resistant to changes in F1 compared to F2. In addition, a pattern that emerges across the significant effects of vowel length on vowel formants is that significant contrast enhancement is observed for peripheral vowels /ɑ e i u/. In other words, the findings here suggest that peripheral vowels peripheralize in long vowels whereas nonperipheral vowels exhibit a lesser degree of vowel length effects on vowel formants.

In summary, although vowel length effects on vowel F1 does not apply to all vowels, the vowels that exhibit F1 movement do not move in one direction but rather move in both directions, effectively making the vowel space larger in F1 for long vowels compared to short vowels. Along with the bidirectional movements in F2 in long vowels, the formant movements in long vowels yield more peripheral vowels (cf. Ünal-Logacev et al., 2019) and a larger vowel space. These results also parallel the findings from disyllabic nonwords in which longer V_1 vowels were more peripheral than shorter V_2 vowels, with greater formant movements in F2 than in F1 and in peripheral vowels than in nonperipheral vowels. As before, I argue that hyperarticulation in F2 for peripheral vowels might serve a contrast enhancement function with respect to vowel backness and rounding.

2.5.2. Vowel fronting

Turkish vowel fronting is described as occurring in the context of both preceding and following tautosyllabic palatal consonantal allophones [ɟ c ʎ] for the back vowels /ɑ o u/, where the palatal allophones are lexically specified in borrowings. In their description of the palatal allophony for /k g l/ and the tautosyllabic vowel fronting that this palatalization conditions, Erguvanlı-Taylan (2015) and Börtlü (2020) do not differentiate between [ɟ c ʎ] in terms of their position in the syllable and claim that all three palatal allophones occur in contexts both preceding and following back vowels /ɑ o u/, which are fronted in these contexts. In contrast, Göksel and Kerslake (2005) observe that [ɟ c] occur only before fronted vowels whereas [ʎ] occurs both preceding and following fronted vowels. In a similar vein, Levi (2001) argues that whereas the palatalized lateral approximant [ʎ] is phonemic, plain and palatal velars are non-contrastive word-finally. Whether /k g l/ following back vowels /ɑ o u/ can be palatalized (and hence the vowel fronted) can be assessed by whether or not vowel harmony selects back vowels in suffixation when these consonants are stem final.

As reviewed above, there is general agreement in the literature that palatalized lateral approximant [ʎ] following fronted back vowels /ɑ o u/ are attested in the Turkish lexicon. Importantly, words with fronted back vowel and coda palatalized lateral approximant [ʎ] sequences in the stem-final syllable make 'irregular stems' that exhibit exceptions to vowel harmony (Canalis & Dikmen, 2020; Kunduracı, 2013). In these words, despite the stem-final vowel being phonologically back, vowel harmony selects front vowel suffixes, such as *kalb*-[i] 'heart-ACC' (**kalb*-[ɯ]), *rol*-[y] 'role-ACC' (**rol*-[u]), and *meşgul*-[y] 'busy-ACC' (**meşgul*-[u]). Compare these stems with stem-final fronted back vowels without a coda consonant or a coda consonant that is not a palatal allophone, such as *sela*-[jɯ] 'call to prayer-ACC' (**sela*-[ji]) and *selam*-[ɯ] 'greeting-ACC' (**selam*-[i]), respectively. In these words, the stem-final vowel is fronted and is preceded by a palatalized lateral approximant [ʎ] that licenses its fronting, and yet this palatalized consonant does not license a front vowel suffix. In light of these observations and following Canalis and Dikmen's (2020) claim that stem-final palatalized consonants select front vowels in suffixation, I argue that if coda velars /k g/ can also be palatalized, they should select front vowel suffixes in vowel fronting contexts.

Similar to the 'irregular stems' with a stem-final coda palatalized lateral approximant [ʎ], vowel harmony irregularity is attributed to words with stem-final fronted back vowel and coda voiceless velar stop /k/ sequences prescriptively: *emlak*-[i] 'estate properties-ACC', *helak*-[i] 'ruin-ACC', *idrak*-[i] 'comprehension-ACC', *istimlak*-[i] 'expropriation-ACC'. Given these stems take front suffixes, arguably the stem-final back vowels are fronted and the stem-final consonant is the palatalized allophone [c]. Alternatively, Clements and Sezer (1982) argue that the stem-final velar stop is plain in this group of words when the stem is bare or a consonant-initial suffix is attached but palatalized when a vowel-initial suffix is attached, in which case the stem-final consonant is no longer tautosyllabic with the stem-final vowel due to resyllabification. However, for this group of words, back vowel suffixes are also acceptable: *emlak*-[ɯ] 'estate properties-ACC', *helak*-[ɯ] 'ruin-ACC', *idrak*-[ɯ] 'comprehension-ACC', *istimlak*-[ɯ] 'expropriation-ACC'. Moreover, Canalis and Dikmen (2020) find that young Turkish speakers do not select front suffixes for the stem *idrak* 'comprehension' (see also Levi, 2004) and do not produce a more palatal voiceless velar stop at the end of the word *idrak* 'comprehension' compared to a near minimal pair *orak* 'grasshook' as evidenced by F2 values at stop release not being significantly different for the two words. As such, Canalis and Dikmen (2020) argue that there is a language change in process whereby palatalized voiceless stop [c] is being lost as an allophone of /k/ in coda position.

In addition to the observations above, note that in the words with purported coda palatalized [c] listed here and in the literature, a majority of the stem-final syllables have a palatalized lateral approximant [ʎ] preceding the vowel. In contrast, in words with stem-final /k/ preceded by fronted /o u/ due to syllable-initial palatalized lateral approximant [ʎ], only back vowel suffixes are allowed: *overlok*-[u] 'overlock-ACC' (**overlok*-[y]), *mahluk*-[u] 'creature-ACC' (**mahluk*-[y]). Thus, although a fronted /o/ or /u/ vowel in these words raises the possibility that the following consonant might palatalize, palatalization of the velar stop is not observed. Consequently, I assume that the stem-final consonant in words with fronted back vowels in the stem-finally syllable is voiceless velar stop /k/ instead of palatalized [c], and the vowel fronting in these words is attributable to the preceding palatalized lateral approximant [ʎ] only.

As for coda voiced velar stop /g/, it is not as frequent as coda /k l/, possibly due to Turkish disfavoring voicing in syllable-final positions. Hence, it is less probable to find irregular vowel harmony behavior with coda /g/ or [ɟ]. One exception might be *alg*-[i] 'algae-ACC' (**alg*-[ɯ]), in which a fronted /ɑ/ is followed by a palatalized lateral approximant [ʎ] before the stem-final palatalized velar stop [ɟ], and hence the front vowel suffix can be accounted for by the palatalized lateral approximant [ʎ] in the coda. There are a relatively large number of stems that end in <log> with a palatalized lateral approximant [ʎ] preceding a fronted /o/, which all behave as regular vowel harmonic stems, such as *blog*-[u] 'blog-ACC' (**blog*-[y]), *diyalog*-[u] 'dialogue-ACC' (**diyalog*-[y]), *psikolog*-[u] 'psychologist-ACC' (**psikolog*-[y]). In this set of words, I assume that the syllable-final consonant is a voiced velar stop [g] instead of a palatalized [ɟ], and that vowel fronting is attributable to the preceding palatalized lateral approximant [ʎ] only.

Together, these observations support Göksel and Kerslake's (2005) description of vowel fronting in Turkish and suggest that the vowel fronting contexts are tautosyllabic preceding [ɟ c ʎ] and following [ʎ]. Accordingly, the vowel fronting (near) minimal pair stimuli here include preceding and following [ʎ] contexts for back vowels /ɑ o u/ and preceding [c ɟ] for back vowels /ɑ o u/ where possible, although I treat palatalized lateral approximant [ʎ] as the main vowel fronting context when there are two competing consonants that might be palatalized and cause vowel fronting such as in *Haluk* (proper name) and *gol* 'goal' (see Table 2.22, note that non-palatalized laterals are velarized in the context of back vowels).

Table 2.22. Minimal or minimal pairs with respect to palatalized consonants in the production stimuli.

Vowel	Context consonant	Target word	Phonetic transcription	Gloss
/ɑ/	/g/ ([ɟ])	Gafur	gɑ.fuɾ	(proper name)
		gavur	ɟa.vuɾ	Non-muslim
	/k/ ([c])	kar	kɑɾ	snow
		kâr	ka:ɾ	profit
	/k/ ([c])	kayıt	kɑ.jɯt	record
		kağıt	ca.ɯt	paper
	/ɫ/ ([ʎ])	lan	ɫɑn	dude (slang)
		lam	ʎam	glass slide
	/ɫ/ ([ʎ])	nal	nɑɫ	horseshoe
		lâl	ʎaʎ	mute
	/ɫ/ ([ʎ])	pedal	pe.dɑɫ	pedal
		metal	me.taʎ	metal
	/ɫ/ ([ʎ])	çuval	t͡ʃuvɑɫ	sack
		tuval	tu.vaʎ	canvas
/o/	/ɫ/ ([ʎ])	gold/kol	goɫd/koɫ	gold/arm
		gol	ɟoʎ	goal
	/ɫ/ ([ʎ])	sol/bol	soɫ/boɫ	left/loose
		sol	soʎ	G (musical note)
	/ɫ/ ([ʎ])	tablo	tɑb.ɫo	painting
		biblo	bib.ʎo	trinket
	/ɫ/ ([ʎ])	tablom	tɑb.ɫom	my painting
		biblom	bib.ʎom	my trinket
	/ɫ/ ([ʎ])	şablon	ʃɑb.ɫon	template
		biblon	bib.ʎon	your trinket
/u/	/k/ ([c])	kuşe	ku.ʃe	glossy paper
		Kûfe	cu:.fe	Kufah (a city in Iraq)
	/k/ ([c])	vuku	vu.ku:	incidence
		rüku	ɾy.cu:	kneel
	/k/ ([c])	kum	kum	sand
		mahkum	mɑh.cum	convict
	/ɫ/ ([ʎ])	usul	u.suɫ	slow
		usul	u.suʎ	method
	/ɫ/ ([ʎ])	kul	kuɫ	servant
		ful	fuʎ	all
	/ɫ/ ([ʎ])	sulu	su.ɫu	runny
		flu	fʎu:	blurred
	/ɫ/ ([ʎ])	suluk	su.ɫuk	water bottle
		Haluk	ha.ʎuk	(proper name)

2.5.2.1. Data analysis and results

There were 2112 vowel tokens in total that make (near) minimal pairs with respect to palatalization of a tautosyllabic context consonant. The phonetic measure of interest was vowel F2 at vowel midpoint, which was modeled as the dependent variable using linear mixed effects models with vowel category (/ɑ/, /o/, or /u/), palatalization of the context consonant, speaker gender, and the interaction of vowel category and palatalization as predictors. As not all vowel categories that are subject to fronting are attested with all consonant types that are subject to palatalization, consonant category was not included as a predictor. There were random intercepts for speaker and minimal pairs, and there was a random slope for palatalization within each minimal pair (11). Multi-level fixed effects were evaluated via a type III ANOVA using Satterthwaite's Method. Vowel category was sum coded. Palatalization had two levels (non-palatalized and palatalized) and was sum coded for ease of interpretation (non-palatalized: -1 and palatalized: 1). Planned comparisons between levels of a variable were performed using the `emmeans` package (Lenth & Piaskowski, 2025).

(11) vowel F2 ~ vowel category
+ palatalization
+ speaker gender
+ vowel category : palatalization
+ (1 | speaker)
+ (1 + palatalization | minimal pair)

There was a significant main effect of vowel category ($F(2, 18.99) = 11.91$, $p = 0.0004$), a significant main effect of palatalization ($\beta = 51.04$, $SE = 14.21$, $t = 3.59$, $p = 0.002$), and a significant main effect of speaker gender ($\beta = 61.66$, $SE = 10.64$, $t = 5.79$, $p < 0.0001$; Table 2.23), indicating that female speakers had higher overall vowel F2 than male speakers. The significant main effect of vowel category indicates that when controlling for palatalization, vowel F2 varied by vowel category. The significant main effect of palatalization indicates that across the vowel categories, vowels in palatalized contexts had significantly higher F2 compared to vowels in non-palatalized contexts. There was no significant interaction of vowel category and palatalization ($F(2, 19.03) = 0.57$, $p = 0.58$), indicating that the difference in F2 between vowels in palatalized and non-palatalized contexts were similar for all vowel categories. Planned comparisons revealed that the 106.4 Hz increase in F2 in /ɑ/ vowels in palatalized contexts (SE = 50.4, $t = 2.109$, $p = 0.046$) and the 138.0 Hz increase in F2 in /u/ vowels in palatalized contexts were significant (SE = 50.4, $t = 2.74$, $p = 0.012$), whereas the 61.8 Hz increase in F2 in /o/ vowels in palatalized contexts was not significant (SE = 59.5, $t = 2.74$, $p = 0.31$; Figure 2.22; Figure 2.23).

Table 2.23. Model results for vowel fronting at vowel midpoint.

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	1193.35	29.22	23.45	40.83	< 0.0001***
/ɑ/	169.48	37.87	19.00	4.47	0.0003***
/o/	-168.01	41.25	18.96	-4.07	0.0007***
Palatalization	51.04	14.21	19.03	3.59	0.002**
Speaker gender	61.66	10.64	25.23	5.79	< 0.0001***
/ɑ/: Palatalization	2.16	19.51	19.09	0.11	0.91
/o/: Palatalization	-20.12	21.22	18.95	-0.95	0.35

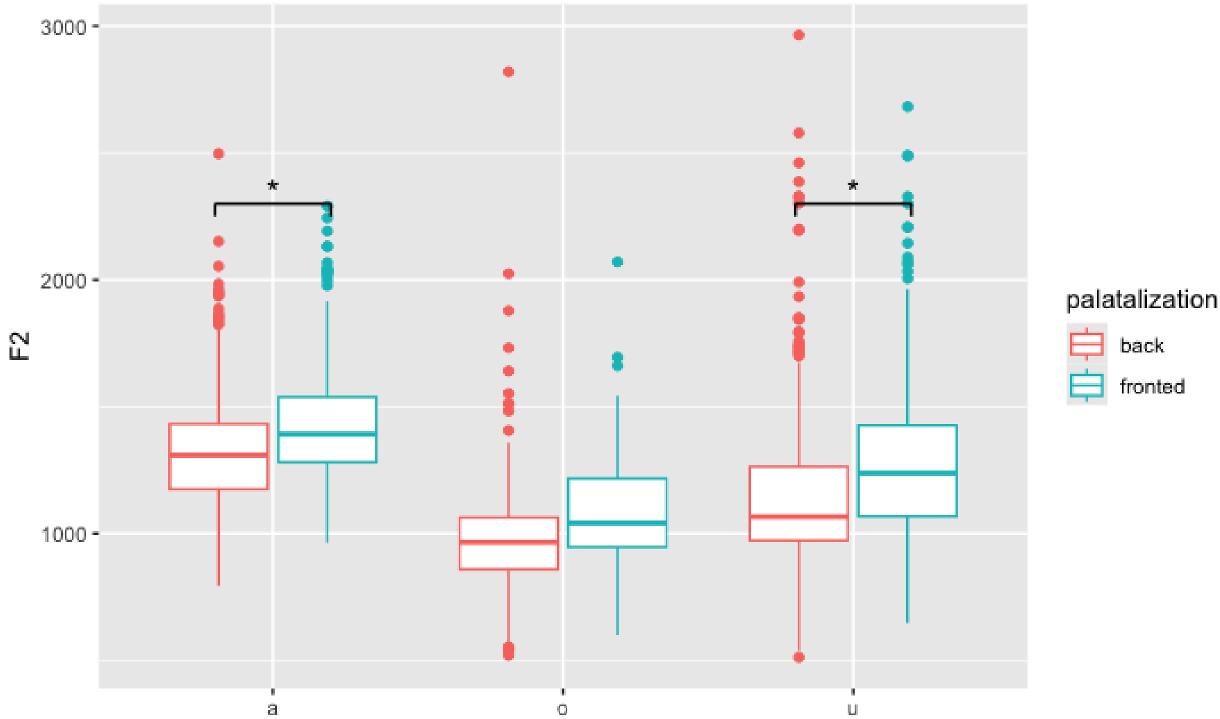

Figure 2.22. Vowel fronting in palatalized consonant contexts.

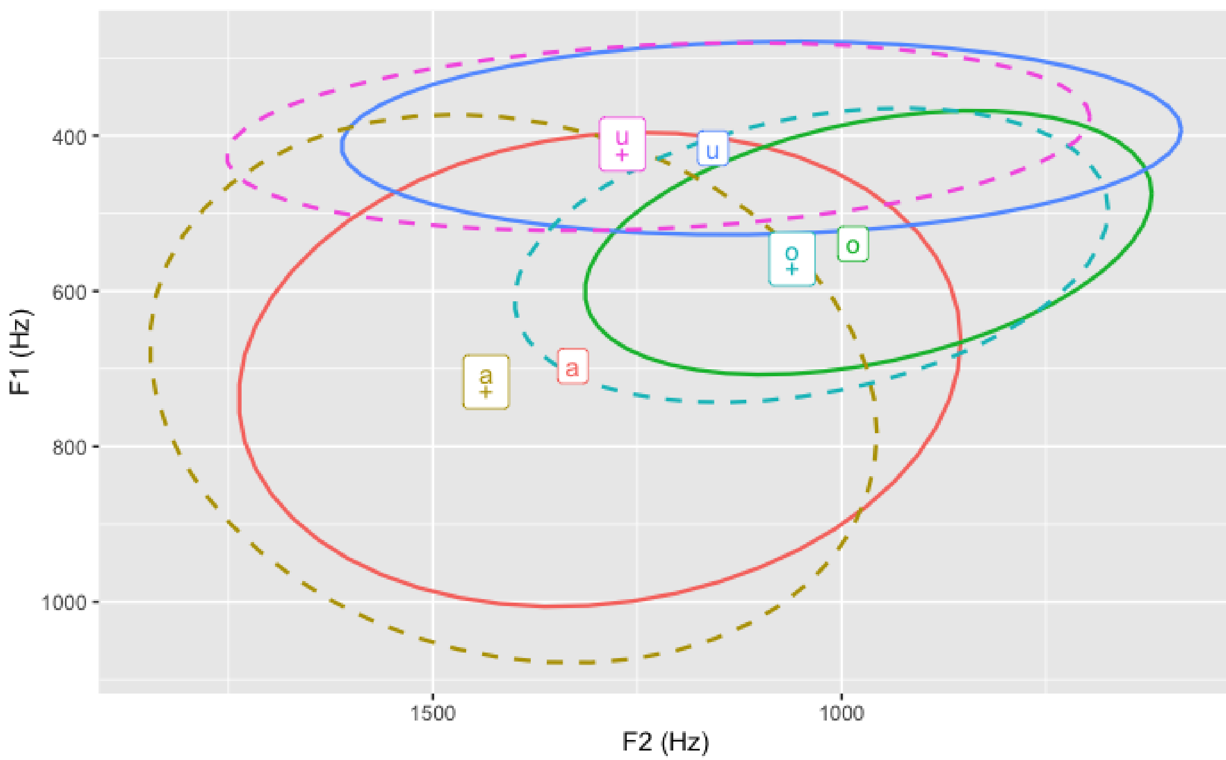

Figure 2.23. Vowel ellipses depicting back vowels (solid ellipses) and fronted vowels (dashed ellipses) in the (near) minimal pair words with respect to palatalization of the context consonant.

Following Canalis and Dikmen (2020), I also ran separate models for preceding and following palatal consonant contexts with vowel F2 measured at vowel onset (20%) and vowel offset (80%), respectively, as the dependent variable. For F2 at vowel onset, there was a significant main effect of palatalization ($\beta = 131.26$, $SE = 20.67$, $t = 6.35$, $p < 0.0001$) indicating increased F2 when the consonant preceding the vowel is palatalized compared to when it is non-palatalized. There was no significant interaction of palatalization and vowel category ($F(2, 13.15) = 1.11$, $p = 0.36$), indicating that palatalization has a similar effect across the vowel categories. Planned comparisons revealed significant increases in F2 at vowel onset for all three vowels preceded by a palatalized consonant (/ɑ/: $\Delta F2 = 207$, $SE = 73.7$, $t = 2.81$, $p = 0.01$; /o/: $\Delta F2 = 241$, $SE = 95.4$, $t = 2.52$, $p = 0.02$; /u/: $\Delta F2 = 340$, $SE = 73.7$, $t = 4.61$, $p = 0.0003$). For F2 at vowel offset, there was a significant main effect of palatalization ($\beta = 101.73$, $SE = 12.28$, $t = 8.37$, $p = 0.0002$). There was also a significant interaction of palatalization and vowel category ($F(2, 5.64) = 12.41$, $p = 0.009$), indicating that the effect of palatalization on vowel F2 varied across the vowel categories. Planned comparisons revealed that /ɑ/ and /u/ exhibited significant increases in F2 when followed by a palatalized consonant ($\Delta F2 = 377.7$, $SE = 62.0$, $t = 6.09$, $p < 0.0001$; $\Delta F2 = 140.0$, $SE = 62.0$, $t = 2.26$, $p = 0.04$, respectively) whereas /o/ exhibited a nonsignificant increase in F2 when followed by a palatalized consonant ($\Delta F2 = 98.7$, $SE = 59.6$, $t = 1.66$, $p = 0.12$). Overall, the results suggest larger increases in F2 at vowel onset or offset than at vowel midpoint, supporting Canalis and Dikmen's (2020) claim that back vowel fronting exhibits consonant-to-vowel coarticulation patterns.

2.5.2.2. Discussion

The results here show that in the context of palatalized consonants [ɟ c ʎ], back vowels /ɑ o u/ are fronted. As such, these results confirm Erguvanlı-Taylan's (2015) and Göksel and Kerslake's (2005) descriptions and Börtlü's (2020) observations regarding back vowel fronting in Turkish. However, note that planned comparisons also show that /o/ fronting does not result in significantly higher F2 at vowel midpoint and vowel offset, despite /o/ being phonetically more back than /ɑ/ and /u/ (in the context of non-palatalized consonants, estimated mean F2 of /o/ is 994 Hz whereas estimated mean F2s of /ɑ/ and /u/ are 1310 Hz and 1123 Hz, respectively). The non-significant increase in F2 in fronted /o/ vowels is accompanied by the fact that the distribution of fronted /o/ is limited to preceding and following palatalized [ʎ] unlike the distributions of fronted /ɑ/ and /u/, suggesting that /o/ might be resistant to fronting.

Arguably, that /o/ and its front counterpart /œ/ are neighboring each other in the vowel space might limit how much /o/ can front before being confused with /œ/ or being ambiguous in its backness. In this regard, /o/ is more similar to back vowel /ɯ/, which is not attested in palatalized consonant contexts in the Turkish lexicon. Similar to /o/, the reason why /ɯ/ does not undergo fronting might be its more ambiguous phonetic backness due to its more central than back realization (see Kılıç & Öğüt, 2004; Radisic, 2014). In contrast to /ɯ/ and /o/, the back vowels /ɑ/ and /u/ front in all possible fronting contexts, arguably because /ɑ/ and /u/ and their front vowel counterparts /e/ and /y/, respectively, are phonetically more distant. For the vowel pair /ɑ e/, the contrast involves both backness and vowel height, as /ɑ/ is back and phonetically low and /e/ is front and phonetically mid. On the other hand, the vowel pair /u y/ are not phonetic neighbors as the back unrounded high vowel /ɯ/ intervenes between the two vowels. The relatively greater phonetic distance between the vowel pairs /ɑ e/ and /u y/ might help contrast preservation even when /ɑ/ and /u/ are fronted. Hence, I argue that vowel fronting in Turkish back vowels might be limited by motivations relating to contrast preservation to avoid neighboring vowels in the vowel space being confused with each other (cf. Bellik, 2018).

2.6. Vowel lowering

As reviewed above, Turkish vowel allophones as discussed in the literature are conditioned by the coda context, with the exception of vowel fronting, which can be due to either preceding or following tautosyllabic consonantal context. The coda contexts that are claimed to condition allophonic variation in vowels are nasal codas /m n/, sonorant consonant codas /l m n r/, and word-final zero codas /∅/. In this section, I present a phonetic analysis of F1 and F2 of the 8 Turkish vowels in each of these contexts in monosyllabic Turkish words, where available, compared to voiceless bilabial obstruent /p/ coda context in monosyllabic words as the baseline.

The phonetic measures of interest were vowel F1 and vowel F2 at vowel midpoint. Each of these measures were modeled as the dependent variable in separate linear mixed effects models with vowel category, coda context, speaker gender, and the interaction of vowel category and coda context as predictors. There were random intercepts for speaker and word (12). Vowel category was sum coded. Coda context had 6 levels (/p/, /l/, /m/, /n/, /r/, /∅/), was treatment coded, and /p/ was selected as the baseline level. Coda contexts were treated as individual levels rather than by their phonological features or classes to reveal potential differences among consonants within a phonological class. Fixed effects with more than 2 levels were evaluated via a type III ANOVA using Satterthwaite's Method. Planned comparisons between levels of a variable were performed using the `emmeans` package (Lenth & Piaskowski, 2025) and Bonferroni adjustment.

(12) vowel formant~ vowel category

+ coda context

+ speaker gender

+ vowel category : coda context

+ (1 | speaker)

+ (1 | item)

Due to the vowels /œ/ not being attested in word-final open syllables in Turkish words, I examined vowel lowering in separate models with and without /œ/ and /Ø/ contexts (Table 2.24). Models 1 and 2 examine the 7 Turkish vowels excluding /œ/ in word-final open syllables of monosyllabic or disyllabic words in addition to nasal and sonorant coda contexts in monosyllabic words. Models 3 and 4 examine the 8 Turkish vowels including /œ/ in monosyllabic words with nasal codas and sonorant consonant codas as compared to voiceless bilabial obstruent coda context as the baseline. The zero coda context is not included in this model as the data does not include observations for the vowel /œ/ in the zero coda context. There is a total of 4894 vowel tokens examined in Models 1 and 2 and a total of 4016 vowel tokens examined in Models 3 and 4.

Table 2.24. Models examining vowel lowering in Turkish.

		Levels of vowel category		Levels of coda context		
Model	**Dependent variable**	**/ɑ e ɯ i o u y/**	**/œ/**	**Sonorant coda**	**/Ø/ coda**	**/p/ coda**
Model 1	Vowel F1	Yes	No	Yes	Yes	Baseline
Model 2	Vowel F2	Yes	No	Yes	Yes	Baseline
Model 3	Vowel F1	Yes	Yes	Yes	No	Baseline
Model 4	Vowel F2	Yes	Yes	Yes	No	Baseline

In Model 1, the dependent variable was vowel F1 at vowel midpoint. There was a significant simple effect of vowel category ($F(6, 83.76) = 1051.51$, $p < 0.0001$), a significant main effect of coda context ($F(5, 89.28) = 30.47$, $p < 0.0001$), and a significant effect of speaker gender ($\beta = 21.07$, $SE = 3.46$, $t = 6.08$, $p < 0.0001$), indicating that vowel F1 differed by vowel category, coda context, and speaker gender. Across the vowel categories, vowel F1 was significantly higher in /l m n ɾ/ contexts compared to /p/ contexts ($\beta = 28.23$, $SE = 7.17$, $t = 3.94$, $p = 0.0002$; /m/: $\beta = 45.21$, $SE = 7.49$, $t = 6.04$, $p < 0.0001$; /n/: $\beta = 59.49$, $SE = 7.64$, $t = 7.78$, $p < 0.0001$; /ɾ/: $\beta = 52.36$, $SE = 7.64$, $t = 6.85$, $p < 0.0001$). In contrast, vowel F1 was not significantly different in word-final open syllable contexts compared to /p/ contexts ($\beta = -6.02$, $SE = 7.13$, $t = -0.85$, $p = 0.4$). There was also a significant interaction of vowel category and coda context ($F(30, 82.51) = 19.62$, $p < 0.0001$), indicating that the effect of coda context on vowel F1 varied across the vowel categories (Table 2.25).

Table 2.25. Model 1 results.

	Estimate	SE	t	p
(Intercept)	482.94	10.94	44.14	< .0001***
Coda /l/	28.23	7.17	3.94	.0002***
Coda /m/	45.21	7.49	6.04	< .0001***
Coda /n/	59.49	7.64	7.78	< .0001***
Coda /ɾ/	52.36	7.64	6.85	< .0001***
Coda /∅/	-6.02	7.12	-0.84	.4005
/ɑ/	262.3	10.56	24.84	< .0001***
/e/	77.32	13.92	5.56	< .0001***
/i/	-124.61	13.92	-8.96	< .0001***
/o/	55.48	10.56	5.25	< .0001***
/u/	-87.9	13.92	-6.32	< .0001***
/ɯ/	-80.44	13.92	-5.78	< .0001***
Speaker gender	21.07	3.46	6.08	< .0001***
/ɑ/: Coda /l/	-42.17	13.39	-3.15	.0024**
/ɑ/: Coda /m/	-25.13	14.15	-1.78	.0801
/ɑ/: Coda /n/	-16.47	13.65	-1.21	.2317
/ɑ/: Coda /ɾ/	-27.66	13.64	-2.03	.0464*
/ɑ/: Coda /∅/	-82.59	15.09	-5.47	< .0001***
/e/: Coda /l/	187.63	17.61	10.66	< .0001***
/e/: Coda /m/	155.25	17.74	8.75	< .0001***
/e/: Coda /n/	192.44	17.82	10.8	< .0001***
/e/: Coda /ɾ/	180.26	20.35	8.86	< .0001***
/e/: Coda /∅/	-7.02	15.9	-0.44	.6603
/i/: Coda /l/	-35.08	17.61	-1.99	.0502
/i/: Coda /m/	-28.71	17.74	-1.62	.1101
/i/: Coda /n/	-43.69	17.8	-2.45	.0166*
/i/: Coda /ɾ/	-3.63	17.82	-0.2	.8391
/i/: Coda /∅/	55.22	22.5	2.45	.0159*
/o/: Coda /l/	-48.4	13.37	-3.62	.0006***
/o/: Coda /m/	-20.68	18.16	-1.14	.2583
/o/: Coda /n/	17.1	18.22	0.94	.3510
/o/: Coda /ɾ/	-65.23	15.33	-4.26	< .0001***
/o/: Coda /∅/	-45.46	15.09	-3.01	.0035**
/u/: Coda /l/	-22.92	17.56	-1.3	.1961
/u/: Coda /m/	2.73	17.74	0.15	.8782
/u/: Coda /n/	-49.19	17.8	-2.76	.0073**
/u/: Coda /ɾ/	-24.5	17.8	-1.38	.1731
/u/: Coda /∅/	11.42	15.55	0.73	.4651
/ɯ/: Coda /l/	5.59	17.61	0.32	.7517
/ɯ/: Coda /m/	-26.59	17.74	-1.5	.1383
/ɯ/: Coda /n/	-22.68	20.35	-1.11	.2687
/ɯ/: Coda /ɾ/	-15.48	20.35	-0.76	.4492
/ɯ/: Coda /∅/	42.37	16.2	2.62	.0109*

The significant interaction of vowel category and coda context was mainly driven by the vowel /e/ (Figure 2.24). Planned comparisons revealed that, compared to baseline /p/ coda context, /e/ vowels preceding lateral /l/ codas had 215.86 Hz higher F1 (SE = 19.3, z = 11.18, $p <$ 0.0001), /e/ vowels preceding bilabial nasal /m/ codas had 200.47 Hz higher F1 (SE = 19.3, z = 10.38, $p < 0.0001$), /e/ vowels preceding nasal /n/ codas had 251.93 Hz higher F1 (SE = 19.3, z = 13.04, $p < 0.0001$), and /e/ vowels preceding /ɾ/ codas had 232.62 Hz higher F1 (SE = 19.3, z = 10.31, $p < 0.0001$). However, /e/ vowels preceding zero coda contexts did not have significantly different F1 compared to baseline /p/ coda context ($p = 1$). These results suggest that /e/ vowels preceding sonorant codas as a class have higher F1 compared to /e/ vowels preceding /p/ codas. However, my results do not support Dadan et al.'s (2024) finding that in word-final open syllables, /e/ surfaces as a lower allophone that is distinct from the allophone preceding sonorant codas.

Several of the planned comparisons revealed significantly lower F1 in the zero coda consonant context compared to baseline /p/ codas. For the vowel /ɑ/, zero codas yielded 88.61 Hz lower F1 than /p/ codas (SE = 15.9, z = -5.59, $p < 0.0001$). Similarly, for the vowel /o/, zero codas yielded 51.48 Hz lower F1 than /p/ codas (SE = 15.9, z = -3.25, p = 0.006), and /n/ codas yielded 76.58 Hz higher F1 than /p/ codas (SE = 19.7, z = 3.89, p = 0.0005). None of the remaining planned comparisons for the vowels /ɯ i œ u y/ were significant, suggesting that these vowels do not exhibit changes in vowel F1 based on coda contexts.

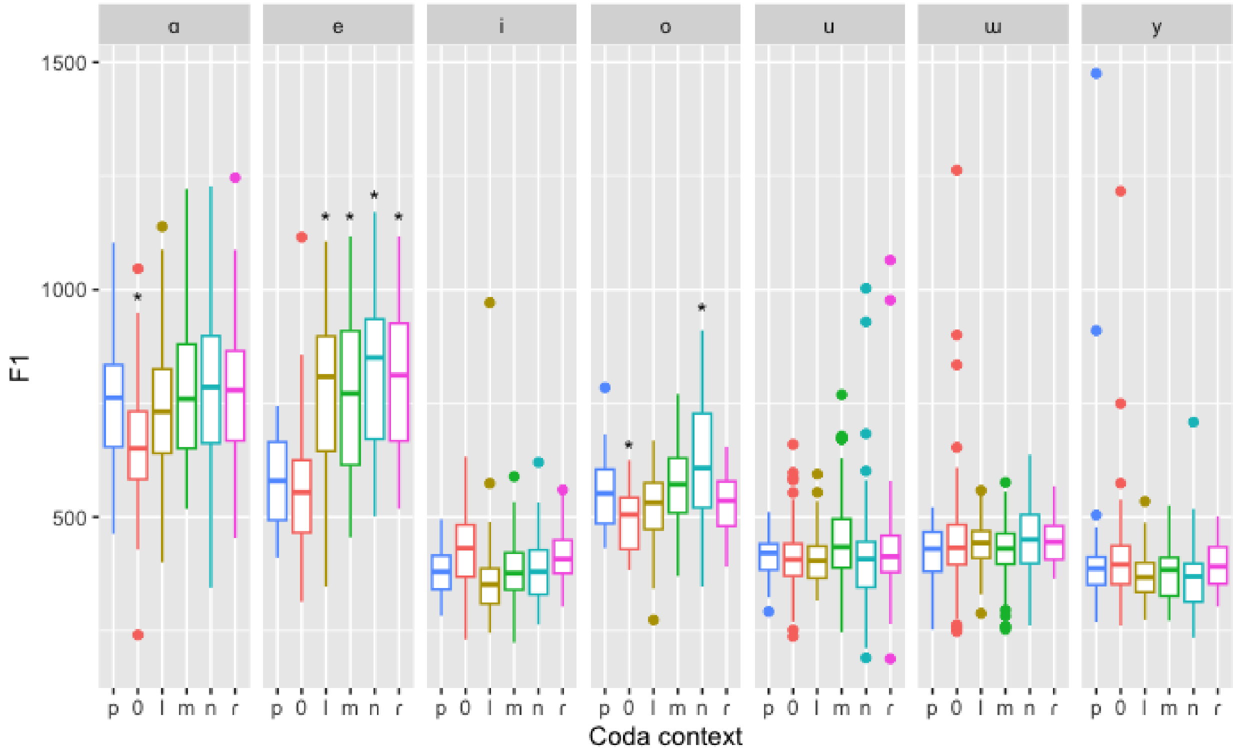

Figure 2.24. Pairwise comparisons of vowel F1 at vowel midpoint across the coda contexts /∅ l m n ɾ / and the reference coda context /p/ for the 7 Turkish vowels.

In Model 2, the dependent variable was vowel F2 at vowel midpoint. There was a significant simple effect of vowel category ($F(6, 86.00) = 339.41$, $p < 0.0001$), and a significant main effect of coda context ($F(5, 87.56) = 7.33$, $p < 0.0001$), indicating that vowel F2 differed by vowel category and coda context. Compared to coda /p/, the coda contexts /n ∅/ significantly increased vowel F2 across the vowel categories ($\beta = 72.04$, $SE = 31.85$, $t = 2.26$, $p = 0.03$; $\beta = 95.13$, $SE = 29.32$, $t = 3.25$, $p = 0.002$, respectively). There was also a significant interaction of vowel category and coda context ($F(30, 86.10) = 4.22$, $p < 0.0001$), indicating that vowel F2 across the coda contexts differed across the vowel categories (Table 2.26).

Table 2.26. Model 2 results.

	Estimate	SE	t	p
(Intercept)	1463.57	30.10	48.63	< 0.0001***
Coda /l/	-38.07	29.92	-1.27	.21
Coda /m/	-5.15	31.22	-0.17	.87
Coda /n/	72.04	31.85	2.26	0.026*
Coda /ɾ/	15.04	31.85	0.47	0.64
Coda /∅/	95.13	29.32	3.25	0.0017**
/ɑ/	-199.79	46.24	-4.32	< 0.0001***
/e/	411.88	60.90	6.76	< 0.0001***
/i/	691.13	60.90	11.35	< 0.0001***
/o/	-488.06	46.24	-10.56	< 0.0001***
/u/	-448.27	60.90	-7.36	< 0.0001***
/ɯ/	-140.64	60.90	-2.31	0.024*
Speaker gender	10.27	8.03	1.28	0.20
/ɑ/: Coda /l/	-43.99	57.46	-0.77	0.44
/ɑ/: Coda /m/	-22.84	60.49	-0.38	0.71
/ɑ/: Coda /n/	-120.90	58.49	-2.07	0.042*
/ɑ/: Coda /ɾ/	11.98	58.47	0.21	0.83
/ɑ/: Coda /∅/	31.19	64.04	0.49	0.63
/e/: Coda /l/	-207.50	75.53	-2.75	0.007**
/e/: Coda /m/	-141.67	76.06	-1.86	0.066
/e/: Coda /n/	-334.38	76.33	-4.38	< 0.0001***
/e/: Coda /ɾ/	-223.06	86.53	-2.58	0.012*
/e/: Coda /∅/	-88.80	68.41	-1.30	0.20
/i/: Coda /l/	121.72	75.53	1.61	0.11
/i/: Coda /m/	25.62	76.06	0.34	0.74
/i/: Coda /n/	94.93	76.32	1.25	0.22
/i/: Coda /ɾ/	-112.53	76.33	-1.48	0.14
/i/: Coda /∅/	-310.78	87.62	-3.55	0.0006***
/o/: Coda /l/	-39.04	57.43	-0.68	0.50
/o/: Coda /m/	-57.58	76.66	-0.75	0.46
/o/: Coda /n/	-33.71	76.92	-0.44	0.66
/o/: Coda /ɾ/	-16.98	65.21	-0.26	0.80
/o/: Coda /∅/	189.26	64.04	2.96	0.004**
/u/: Coda /l/	12.96	75.48	0.17	0.86
/u/: Coda /m/	129.63	76.06	1.70	0.09
/u/: Coda /n/	13.69	76.32	0.18	0.86
/u/: Coda /ɾ/	114.56	76.32	1.50	0.14
/u/: Coda /∅/	87.22	67.01	1.30	0.20
/ɯ/: Coda /l/	63.92	75.53	0.85	0.40
/ɯ/: Coda /m/	-63.00	76.06	-0.83	0.41
/ɯ/: Coda /n/	188.30	86.53	2.18	0.03*
/ɯ/: Coda /ɾ/	288.51	86.53	3.33	0.001**
/ɯ/: Coda /∅/	147.90	69.62	2.12	0.037*

The significant interaction of vowel category and coda context was mainly driven by the vowel /e/ (Figure 2.25). Planned comparisons revealed significant differences in F2 between baseline and /l n/ consonant contexts. Compared to baseline /p/ coda context, /e/ vowels preceding lateral approximant /l/ codas had 245.57 Hz lower F2 (*SE* = 82.4, *z* = -2.98, *p* = 0.01), and /e/ vowels preceding nasal stop /n/ codas had 262.35 Hz lower F2 (*SE* = 82.4, *z* = -3.18, *p* = 0.007). These results suggest that sonorant codas do not have an effect on F2 of /e/ vowels as a class compared to /p/ codas.

Several of the planned comparisons revealed significantly higher F2 in one or more coda consonant context compared to baseline /p/ codas. For the vowel /o/, /∅/ codas yielded 284.39 Hz higher F2 than /p/ codas (*SE* = 67.5, *z* = 4.21, *p* = 0.0001). For the vowel /y/, /n/ codas yielded 265.69 Hz higher F2 than /p/ codas (*SE* = 95.5, *z* = 2.70, *p* = 0.04). For the vowel /ɯ/, /∅/ codas yielded 243.03 Hz higher F2 than /p/ codas (*SE* = 75.0, *z* = 3.24, *p* = 0.006) and /n/ codas yielded 260.34 Hz higher F2 than /p/ codas (*SE* = 95.5, *z* = 2.73, *p* = 0.03). None of the planned comparisons for the vowels /ɑ i u/ were significant, suggesting that these vowels do not exhibit changes in vowel F2 based on coda consonant.

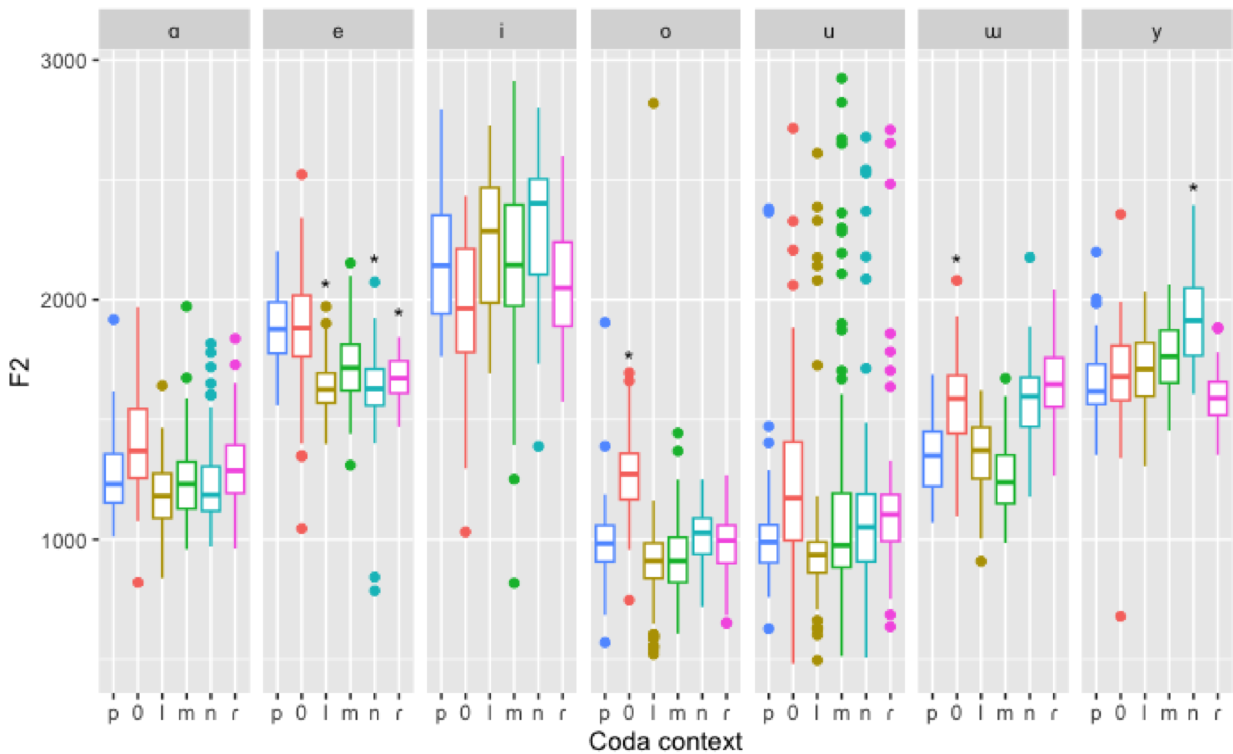

Figure 2.25. Pairwise comparisons of vowel F2 at vowel midpoint across the coda contexts /p Ø l m n ɾ / for the 7 Turkish vowels.

In Model 3, the dependent variable was vowel F1 at vowel midpoint. There was a significant main effect of vowel category (F(7, 62.73) = 1404.09, p < 0.0001), a significant main effect of coda context (F(4, 69.62) = 30.61, p < 0.0001), and a significant main effect of speaker gender (β = 21.57, SE = 3.33, t = 6.48, p < 0.0001), indicating that vowel F1 differed by vowel category, coda context, and speaker gender. Coda /l m n ɾ/ contexts yielded significantly higher F1 compared to /p/ coda contexts across the vowel categories (all ps < 0.0001, see Table 2.27). There was also a significant interaction of vowel category and coda context (F(28, 62.29) = 11.97, p < 0.0001), indicating that vowel F1 across the coda contexts differed across the vowel categories.

Table 2.27. Model 3 results.

	Estimate	SE	t	p
(Intercept)	489.44	11.41	42.89	< 0.0001***
Coda /l/	23.54	5.45	4.32	0.0001***
Coda /m/	43.10	5.66	7.62	< 0.0001***
Coda /n/	55.76	5.76	9.68	< 0.0001***
Coda /ɾ/	47.38	5.76	8.23	< 0.0001***
/ɑ/	254.15	8.02	31.70	< 0.0001***
/e/	70.32	10.81	6.51	< 0.0001***
/i/	-131.61	10.81	-12.17	< 0.0001***
/o/	48.26	8.02	6.02	< 0.0001***
/œ/	50.34	8.49	5.93	< 0.0001***
/u/	-94.90	10.81	-8.78	< 0.0001***
/ɯ/	-87.44	10.81	-8.09	< 0.0001***
Speaker gender	21.57	3.33	6.48	< 0.0001***
/ɑ/: Coda /l/	-35.99	10.48	-3.43	0.0012**
/ɑ/: Coda /m/	-21.52	11.10	-1.94	0.058
/ɑ/: Coda /n/	-11.26	10.65	-1.06	0.30
/ɑ/: Coda /ɾ/	-21.41	10.63	-2.02	0.049*
/e/: Coda /l/	192.66	13.99	13.78	< 0.0001***
/e/: Coda /m/	157.72	14.07	11.21	< 0.0001***
/e/: Coda /n/	196.46	14.13	13.91	< 0.0001***
/e/: Coda /ɾ/	185.58	16.29	11.39	< 0.0001***
/i/: Coda /l/	-30.05	13.99	-2.15	0.036*
/i/: Coda /m/	-26.25	14.07	-1.87	0.067
/i/: Coda /n/	-39.62	14.11	-2.81	0.007**
/i/: Coda /ɾ/	1.61	14.12	0.11	0.91
/o/: Coda /l/	-43.33	10.45	-4.15	0.0001***
/o/: Coda /m/	-17.99	14.55	-1.24	0.22
/o/: Coda /n/	21.39	14.59	1.47	0.15
/o/: Coda /ɾ/	-59.69	12.10	-4.93	< 0.0001***
/œ/: Coda /l/	-36.22	14.73	-2.46	0.017*
/œ/: Coda /m/	-18.59	14.81	-1.26	0.21
/œ/: Coda /n/	-29.80	14.85	-2.01	0.049*
/œ/: Coda /ɾ/	-38.35	14.85	-2.58	0.01*
/u/: Coda /l/	-18.08	13.93	-1.30	0.20
/u/: Coda /m/	5.19	14.07	0.37	0.71
/u/: Coda /n/	-45.12	14.11	-3.20	0.002**
/u/: Coda /ɾ/	-19.17	14.11	-1.36	0.18
/ɯ/: Coda /l/	10.63	13.99	0.76	0.45
/ɯ/: Coda /m/	-24.13	14.07	-1.72	0.09
/ɯ/: Coda /n/	-18.61	16.29	-1.14	0.26
/ɯ/: Coda /ɾ/	-10.15	16.29	-0.62	0.54

The significant interaction of vowel category and coda context was mainly driven by the vowel /e/ (Figure 2.26). Planned comparisons revealed that, compared to baseline /p/ coda context, /e/ vowels preceding lateral /l/ codas had 216.20 Hz higher F1 ($SE = 15.2$, $z = 14.24$, $p < 0.0001$), /e/ vowels preceding bilabial nasal /m/ codas had 200.81 Hz higher F1 ($SE = 15.2$, $z = 13.22$, $p < 0.0001$), /e/ vowels preceding nasal /n/ codas had 252.22 Hz higher F1 ($SE = 15.2$, $z = 16.59$, $p < 0.0001$), and /e/ vowels preceding /ɾ/ codas had 232.96 Hz higher F1 ($SE = 17.9$, $z = 13.04$, $p < 0.0001$). These results suggest that /e/ vowels preceding sonorant codas as a class have higher F1 compared to /e/ vowels preceding /p/ codas.

Several of the planned comparisons revealed significantly higher F1 in one consonant context compared to baseline /p/ codas. For the vowel /ɑ/, /n/ codas yielded 44.51 Hz higher F1 than /p/ codas ($SE = 10.5$, $z = 4.23$, $p = 0.0001$), and the 25.97 Hz difference in F2 between /ɾ/ and /p/ codas was approaching significance ($SE = 10.5$, $z = 2.47$, $p = 0.054$). Similarly, for the vowel /o/, /n/ codas yielded 77.15 Hz higher F1 than /p/ codas ($SE = 15.6$, $z = 4.95$, $p < 0.0001$). For the vowel /i/, /ɾ/ codas yielded 48.99 Hz higher F1 than /p/ codas ($SE = 15.2$, $z = 3.22$, $p = 0.005$). For the vowel /u/, /m/ codas yielded 48.23 Hz higher F1 than /p/ codas ($SE = 15.2$, $z = 3.18$, $p = 0.006$). None of the planned comparisons for the vowels /ɯ œ y/ were significant, suggesting that the central vowels do not exhibit changes in vowel F1 based on coda consonant.

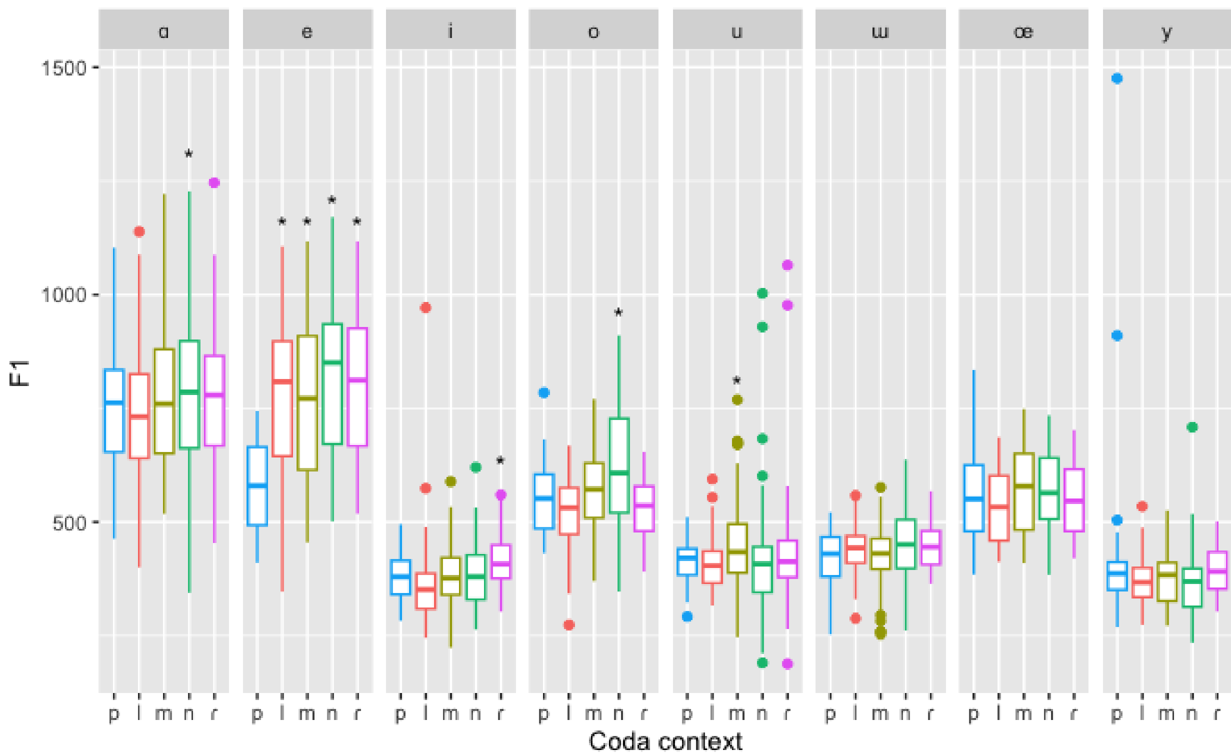

Figure 2.26. Pairwise comparisons of vowel F1 at vowel midpoint across the coda contexts /p l m n ɾ / for the 8 Turkish vowels.

In Model 4, the dependent variable was vowel F2 at vowel midpoint. There was a significant simple effect of vowel category (F(7, 65.11) = 986.84, p < 0.0001), a significant main effect of coda context (F(4, 69.32) = 11.22, p < 0.0001), and a significant main effect of speaker gender (β = 15.65, SE = 7.39, t = 2.12, p = 0.035), indicating that vowel F2 differed by vowel category, coda context, and speaker gender. Coda /l/ contexts had significantly lower F2 whereas coda /n/ contexts had significantly higher F2 compared to /p/ coda contexts across the vowel categories (ps < 0.05, see Table 2.28). There was also a significant interaction of vowel category and coda context (F(28, 62.25) = 9.84, p < 0.0001), indicating that vowel F2 across the coda contexts differed across the vowel categories.

Table 2.28. Model 4 results.

	Estimate	SE	t	p
(Intercept)	1454.69	18.69	77.83	< 0.0001***
Coda /l/	-38.09	15.21	-2.50	0.015*
Coda /m/	0.09	15.80	0.01	0.99
Coda /n/	61.25	16.08	3.81	0.0003***
Coda /ɾ/	2.39	16.08	0.15	0.88
/ɑ/	-198.07	23.55	-8.41	< 0.0001***
/e/	415.37	31.48	13.20	< 0.0001***
/i/	694.63	31.48	22.07	< 0.0001***
/o/	-484.06	23.55	-20.55	< 0.0001***
/œ/	-23.18	24.36	-0.95	0.35
/u/	-444.77	31.48	-14.13	< 0.0001***
/ɯ/	-137.15	31.48	-4.36	< 0.0001***
Speaker gender	15.65	7.39	2.12	0.035*
/ɑ/: Coda /l/	-38.47	30.18	-1.28	0.21
/ɑ/: Coda /m/	-22.57	31.86	-0.71	0.48
/ɑ/: Coda /n/	-104.60	30.63	-3.42	0.001**
/ɑ/: Coda /ɾ/	30.41	30.59	0.99	0.33
/e/: Coda /l/	-203.76	40.07	-5.09	< 0.0001***
/e/: Coda /m/	-143.18	40.30	-3.55	0.0008***
/e/: Coda /n/	-319.81	40.44	-7.91	< 0.0001***
/e/: Coda /ɾ/	-206.68	46.38	-4.46	< 0.0001***
/i/: Coda /l/	125.47	40.07	3.13	0.003**
/i/: Coda /m/	24.12	40.30	0.60	0.55
/i/: Coda /n/	109.45	40.41	2.71	0.009**
/i/: Coda /ɾ/	-95.85	40.44	-2.37	0.02*
/o/: Coda /l/	-35.76	30.12	-1.19	0.24
/o/: Coda /m/	-59.59	41.30	-1.44	0.15
/o/: Coda /n/	-19.69	41.41	-0.48	0.64
/o/: Coda /ɾ/	-1.11	34.60	-0.03	0.98
/œ/: Coda /l/	-27.59	41.55	-0.66	0.51
/œ/: Coda /m/	9.27	41.77	0.22	0.83
/œ/: Coda /n/	-102.96	41.87	-2.46	0.02*
/œ/: Coda /ɾ/	-116.47	41.87	-2.78	0.007**
/u/: Coda /l/	16.74	39.96	0.42	0.68
/u/: Coda /m/	128.12	40.30	3.18	0.002**
/u/: Coda /n/	28.21	40.41	0.70	0.49
/u/: Coda /ɾ/	130.94	40.41	3.24	0.002**
/ɯ/: Coda /l/	67.67	40.07	1.69	0.097
/ɯ/: Coda /m/	-64.51	40.30	-1.60	0.12
/ɯ/: Coda /n/	202.82	46.38	4.37	< 0.0001***
/ɯ/: Coda /ɾ/	304.89	46.38	6.57	< 0.0001***

The significant interaction of vowel category and coda context was mainly driven by the vowel /e/ (Figure 2.27). Planned comparisons revealed significant differences in F2 between baseline and all 4 coda consonant contexts. Compared to baseline /p/ coda context, /e/ vowels preceding lateral /l/ codas had 241.84 Hz lower F2 ($SE = 43.3$, $z = -5.58$, $p < 0.0001$), /e/ vowels preceding bilabial nasal /m/ codas had 143.09 Hz lower F2 ($SE = 43.3$, $z = -3.30$, $p = 0.009$), /e/ vowels preceding nasal /n/ codas had 258.57 Hz lower F2 ($SE = 43.4$, $z = -5.96$, $p < 0.0001$), and /e/ vowels preceding /ɾ/ codas had 204.29 Hz lower F2 ($SE = 50.7$, $z = -4.03$, $p = 0.0002$). These results suggest that /e/ vowels preceding sonorant codas as a class have lower F2 compared to /e/ vowels preceding /p/ codas.

Several of the planned comparisons revealed significantly different F2 in one or more consonant context compared to baseline /p/ codas. For the vowel /i/, /n/ codas yielded 170.70 Hz higher F2 than /p/ codas ($SE = 43.3$, $z = 3.94$, $p = 0.0003$). For the vowel /y/, /m/ codas yielded 128.42 Hz higher F2 than /p/ codas ($SE = 50.7$, $z = 2.53$, $p = 0.045$) and /n/ codas yielded 267.84 Hz higher F2 than /p/ codas ($SE = 50.7$, $z = 5.28$, $p < 0.0001$). For the vowel /œ/, /ɾ/ codas yielded 113.08 Hz lower F2 than /p/ codas ($SE = 44.9$, $z = -2.54$, $p = 0.045$). For the vowel /u/, /m/ codas yielded 128.21 Hz higher F2 than /p/ codas ($SE = 43.3$, $z = 2.96$, $p = 0.01$) and /ɾ/ codas yielded 133.33 Hz higher F2 than /p/ codas ($SE = 50.7$, $z = 3.08$, $p = 0.01$). For the vowel /ɯ/, /n/ codas yielded 264.07 Hz higher F2 than /p/ codas ($SE = 50.7$, $z = 5.21$, $p < 0.0001$) and /ɾ/ codas yielded 307.28 Hz higher F2 than /p/ codas ($SE = 50.7$, $z = 6.06$, $p < 0.001$). For the vowel /ɑ/, /l/ codas yielded 76.55 Hz lower F2 than /p/ codas ($SE = 30.4$, $z = -2.52$, $p = 0.047$). For the vowel /o/, /l/ codas had 73.84 Hz lower F2 than /p/ codas, which was approaching significance ($SE = 43.3$, $z = -2.44$, $p = 0.06$).

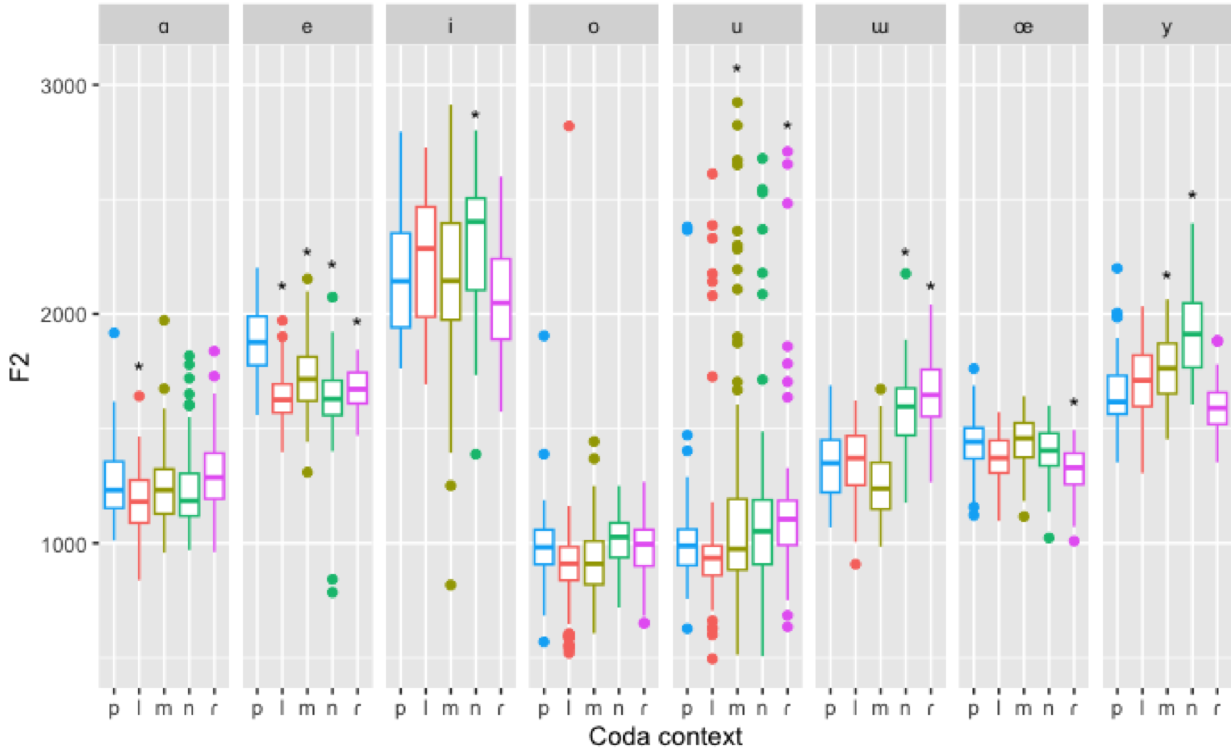

Figure 2.27. Pairwise comparisons of vowel F2 at vowel midpoint across the coda contexts /p l m n ɾ / for the 8 Turkish vowels.

2.6.1. Discussion

Based on descriptive grammars of Turkish, I examined nasal and sonorant coda contexts and word-final open syllable contexts for environments conditioning allophonic variation in Turkish vowels. Whereas these allophonic variations are all described to be in the form of vowel lowering, for /e/, both lowering and backing is predicted. I do not find consistent vowel lowering in the nasal or sonorant coda contexts or in word-final open syllables. In contrast, the results here do support /e/ vowel lowering to [ɛ] in sonorant coda contexts, as sonorant coda consonants yielded significantly higher F1 and lower F2 for the vowel /e/. However, the results here did not support /e/ lowering in word-final open syllables, failing to three replicate the three-way allophonic variation observed in Dadan et al. (2024).

For the other vowels in the various coda contexts, nasal coda consonants exhibited significant effects in a relatively higher number of cases, however these effects did not follow a consistent pattern. The effects of nasal codas manifested in only higher F1 for /ɑ o/, only higher F2 for /ɯ y/, and both higher F1 and F2 for /i u/. Moreover, only /y/ exhibited significant effects when followed by both /m/ and /n/. The nasal consonants yielded higher F1 for all back vowels and higher F2 for all high vowels regardless of vowel backness. However, these effects on F1 were confined to /m/ for high back vowels /ɯ u/ and to /n/ for nonhigh back vowels /ɑ o/. Hence, it is not clear whether nasal coda consonants behave as a class in conditioning the first two vowel formants. However, although oral formant shifts are also documented as acoustic correlates of nasalization in vowels (see Carignan, 2018), it remains possible that nasalization is manifest in other phonetic measures, such as the nasal formants or the amplitude of the first formant, A1 (Chen, 1995), which make nasalized vowels perceptually distinct from their non-nasal counterparts and hence might cue allophony. Nonetheless, given the non-systematic way in which nasal coda consonants manifests in vowel formants, and the even less systematic way in which sonorant coda consonants manifest in vowel formants overall, I do not find evidence in support of allophonic vowel lowering in the context of sonorant coda consonants in Turkish except for lowering of /e/ to [ɛ] (Figure 2.28). Overall, that vowels are resistant movement along F1 is in line with the observations from vowels in Turkish nonwords and words with vowel length contrasts. In contrast, that /e/ can be significantly lowered and backed to [ɛ] suggests that the backness contrast is well preserved in the lower region of the vowel space without causing ambiguity despite phonetic proximity.

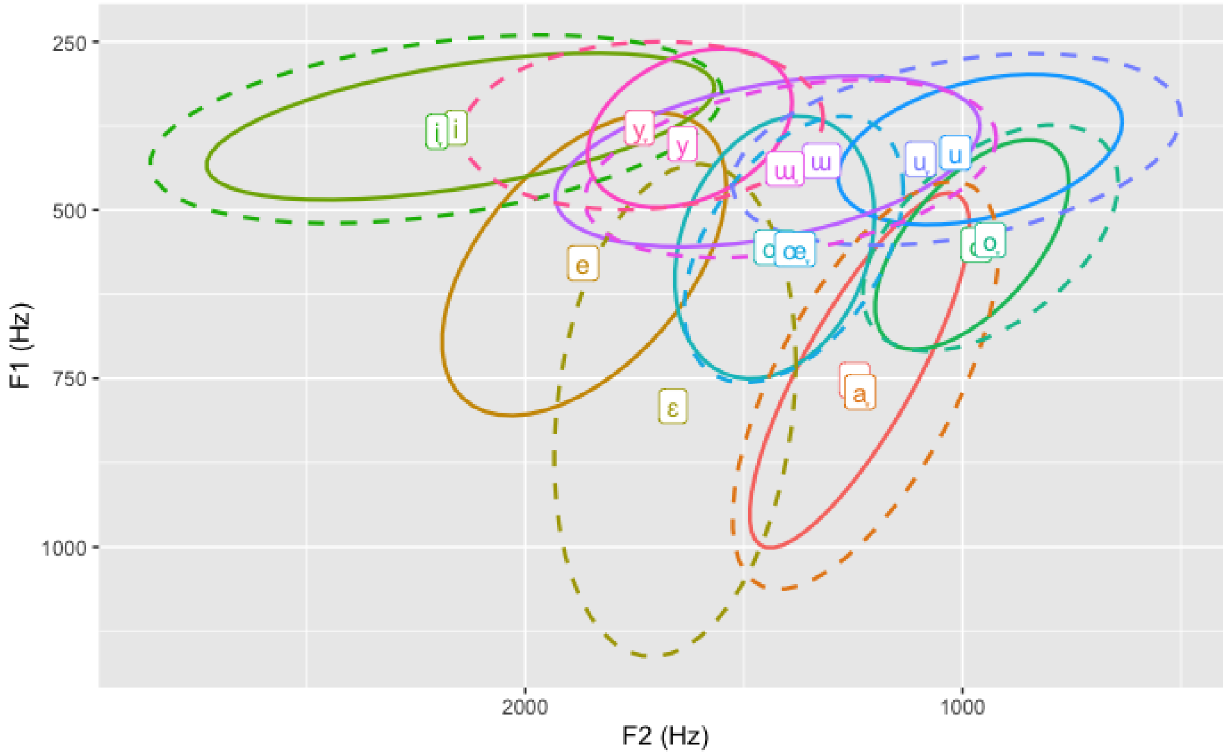

Figure 2.28. Vowel ellipses depicting vowels followed by /p/ (solid ellipses) and vowels followed by sonorant consonants (dashed ellipses).

2.7. General discussion

In this chapter, I analyzed Turkish vowel productions in various phonetic and phonological contexts to explore whether the patterns of variation observed are specific to individual vowels or generalize to harmony classes. I first examined vowel formants in monosyllabic nonwords to capture a description of the Turkish vowel space under a controlled context. I found that Turkish vowels can be classified into three phonetic height categories (high, mid, and low) and yet are continuously distributed along the F2 dimension, suggesting that the front/back and rounded/unrounded harmony classes are not clearly defined in terms of vowel phonetics. I then analyzed vowel formants in disyllabic nonwords to examine phonetic variation due to vowel-to-vowel coarticulation and in Turkish words to examine phonological variation due to vowel length, vowel fronting, and vowel lowering. I did not find evidence suggesting that vowels within a harmony class uniformly participate in coarticulation, hyperarticulation, or other allophonic processes conditioned by contextual consonants. Instead, I found evidence suggesting that peripheral vowels /ɑ i o u/ exhibit more variation overall compared to nonperipheral vowels /ɯ œ y/, and that peripheral vowels are more likely to exhibit hyperarticulation along the F2 dimension. I argued that contrast enhancement in F2 might facilitate the accurate perception of vowel backness and rounding, and that the limited variation observed in the nonperipheral vowels might be attributed to contrast preservation efforts. Together, these results suggest that Turkish harmony classes are defined solely on the basis of vowel cooccurrence and that vowels within a harmony class do not participate in processes that entail variation in vowel formants as a class.

CHAPTER 3

VOWEL DISCRIMINATION EXPERIMENTS

The main research question in this work is whether lexical vowel cooccurrence statistics are one of the factors that influence perceived vowel similarity beyond acoustic-phonetic measures of vowel similarity, especially in vowel harmony languages where vowel cooccurrence is systematic and can lead to the emergence of harmony classes as phonological representations (Cole, 2009). One way in which perceived vowel similarity is measured and the mechanisms and factors behind perceived vowel similarity have been probed is vowel discrimination tasks. In vowel discrimination, ease of discrimination is interpreted as a measure of perceived vowel dissimilarity. That is, the more perceptually similar two vowel stimuli are, the harder it is for listeners to discriminate them. By looking at response times in vowel discrimination, I aim to investigate whether the likelihood of vowel cooccurrence in the lexicon can be one of the phonological predictors of perceived vowel similarity.

One concern for measuring perceived vowel similarity in vowel discrimination tasks to examine the role phonology plays in it is that evidence suggests that vowel discrimination is mainly driven by acoustic-phonetic differences in vowel stimuli rather than vowel phonology, suggesting continuous rather than categorical perception of vowels (e.g., Fry et al., 1962). Hence, part of the undertaking here is to test the hypothesis that vowel discrimination can reflect phonological representations beyond vowel phonetics. In this chapter, I review previous literature that manipulated task demands in discrimination tasks to target either acoustic or phonological processing, and present results from two Turkish vowel discrimination experiments with varying task demands to investigate the relative contributions of acoustic-phonetic and phonological measures of vowel similarity to perceived vowel similarity. I cross-linguistically compare native vowel discrimination by Turkish and English listeners to examine whether vowel cooccurrence patterns in the listeners' native lexicon are reflected in perceived vowel similarity, and whether these lexical statistics are more influential in vowel harmony languages.

3.1. Manipulating task demands in vowel discrimination

Studies that investigate the effects of manipulating the task demands in perceptual discrimination tasks are predominantly interested in the phenomenon termed 'categorical perception' which indicates that the continuous speech signal is processed with respect to speech categories, or alternatively that speech perception involves categorization of the continuous speech signal (McMurray, 2022). Neural evidence suggests that listeners process the speech signal both continuously and categorically, which involve separate neural mechanisms that interact and occur mostly simultaneously but with earlier activation of mechanisms related to continuous processing, suggesting two separate yet interconnected processes (e.g., Toscano et al., 2018; Rizzi & Bidelman, 2024; see Getz & Toscano, 2021 for a review). Similarly, behavioral evidence suggests that phonological and lexical, top-down influences on speech perception become evident later than purely acoustic, bottom-up influences (e.g., Fox, 1984), suggesting that the effects of these two types of information can be measured individually. Paralleling these facts, evidence from vowel and consonant discrimination tasks suggests that task demands can target listening modes that give more weight to either the continuous cues in the speech signal or the category labels of the stimuli (e.g., Pisoni & Tash, 1974; Gerrits & Schouten, 2004; see McMurray, 2022 for a review).

In a relatively early study testing whether 'categorical perception', defined as the inability to perceive the acoustic differences between speech sounds–and specifically consonants–when the stimuli are selected from a single speech category, is a corollary of task demands, Pisoni and

Tash (1974) presented native English listeners with stimulus pairs from a /ba-pa/ continuum in an AX discrimination task. Listeners were instructed to respond with 'same' when they hear a /ba/-/ba/ pair or a /pa/-/pa/ pair and with 'different' when they hear a /ba/-/pa/ pair or a /pa/-/ba/ pair, i.e., respond based on phonemic categorization. There were two types of 'same' stimulus pairs: in 'A-A' pairs an identical stimulus token was presented twice, and in 'A-a' pairs one endpoint stimulus token and a second stimulus token that is two steps from that endpoint token were presented, thus making a 'within category' pair. The 'different' stimulus pairs differed by the number of steps that separated the two across-category stimulus tokens, with stimulus pairs that are 2 steps apart (acoustically closest), 4 steps apart, and 6 steps apart (acoustically most distant). The authors hypothesized that within the 'same' trials, the acoustically same 'A-A' pairs would receive faster 'same' responses than the acoustically different 'A-a' pairs, and that within the 'different' trials, the acoustically more distant stimulus pairs would receive faster 'different' responses than the acoustically closer stimulus pairs. The reasoning behind the hypothesized longer response times for both the acoustically different 'same' trials and the acoustically closer 'different' trials was that the stimulus pairs in these trials would require an additional level of processing involving a feature-level comparison beyond acoustic identity, whereas in acoustically identical 'same' trials and acoustically distant 'different' trials, acoustic and feature-level information are consistent with each other.

Pisoni and Tash's (1974) results were consistent with their predictions, suggesting that finer grained cues are processed even in a consonant discrimination task that requires phonemic categorization, and that task demands can be manipulated to reveal different levels of processing. As such, the authors argue that categorical perception might be a property of the task demands rather than consonant perception itself, with tasks such as ABX discrimination more readily targeting categorical processing (e.g., Liberman et al., 1957) by involving comparisons of the identified category labels of the stimuli and tasks such as AX discrimination more readily targeting continuous processing by involving comparisons of the stimulus acoustics (Pisoni, 1973). This interpretation that ABX discrimination tasks involve categorical processing of the speech stimuli arguably does not apply to vowels, as the evidence suggesting that vowel perception is continuous and not categorical was observed in an ABX vowel discrimination task (Fry et al., 1962). Hence, the challenge for vowel discrimination is to show that vowels can be perceived *categorically* under different task demands that target the categorical listening mode. In this vein, Pisoni (1973) presents a study of AX vowel discrimination where when task demands are manipulated, response patterns suggest relatively more categorical rather than continuous perception of vowels.

Pisoni (1973) hypothesized that continuous perception in vowels (e.g., Fry et al.,1962) as opposed to categorical perception in consonants (e.g., Liberman et al., 1957) indicates differences in how vowels and consonants are encoded by listeners, with phonetic detail being retained in the auditory short-term memory to a higher degree for vowels compared to consonants and listeners relying on phonemic identification in consonant discrimination but on phonetic detail stored in auditory short-term memory in vowel discrimination. To manipulate the degree of phonetic detail being retained in auditory short-term memory, Pisoni compared AX vowel and consonant discrimination with varying lengths of inter-stimulus intervals. Pisoni presented native English listeners with synthetic steady state vowel steps from two /i-ɪ/ continua (with either 300 ms or 50 ms vowel durations) and consonant-vowel sequence steps from /bæ-dæ/ and /ba-pa/ continua in an AX discrimination task. The inter-stimulus intervals were 0 ms, 250 ms, 500 ms, 1000 ms, and 2000 ms.

The results suggested that vowel discrimination accuracy for both across-category and within-category comparisons was highest at 250 ms inter-stimulus interval and dropped with each longer inter-stimulus interval step. In contrast, consonant discrimination accuracy and inter-stimulus interval steps did not have a clear relationship, with overall high accuracy for the between-category comparisons and overall low accuracy for the within-category comparisons. Pisoni (1973) argues that at the 250 ms inter-stimulus interval, vowel discrimination is most continuous and vowels are processed according to their phonetic detail, whereas both shorter and longer inter-stimulus intervals make vowel perception less continuous and more categorical. On the other hand, consonant perception in this task is categorical regardless of the inter-stimulus interval manipulation. Together, these findings from Pisoni (1973) and Pisoni and Tash (1974) suggest that both vowels and consonants can be processed more continuously or more categorically by manipulating task demands.

In related work, Gerrits and Schouten (2004; see also Schouten et al., 2003) claim that the distinction between continuous and categorical perception is whether the task demands require and enable phonemically labeling stimuli; they show that manipulating task demands can influence whether vowel discrimination happens in an auditory listening mode or a labeling mode. Native Dutch listeners were presented with vowel tokens from an /u-i/ continuum in a /p_p/ context in a 4 interval 2 alternative forced choice task (4I2AFC) and a 2 interval 2 alternative forced choice task (2I2AFC). In the 4I2AFC task, stimuli were presented in two orders, ABAA or AABA, and listeners were asked to indicate whether the oddball stimulus occurred as the second or the third stimulus. In the 2I2AFC task, stimuli were presented in two orders, AB or BA (/i/-/u/ or /u/-/i/), and listeners were asked to indicate which order the stimuli were presented in among two response alternatives: “ie-oe” (/i/-/u/) or “oe-ie” (/u/-/i/). Thus, in the 4I2AFC task, listeners had to detect an auditory oddball among three acoustically identical stimuli, and in the 2I2AFC task, listeners were asked to use labels to indicate the order of stimuli. In both tasks, a categorical perception index was calculated based on listeners’ identification responses of the same stimuli, and results suggested more categorical perception of vowels in the 2I2AFC task that required labeling stimuli than in the 4I2AFC task that required detecting auditory dissimilarity. The researchers interpret these results to support a dual processing model whereby listeners process speech both in terms of the acoustic signal and the phoneme labels, and that task demands can target acoustic and phonemic (or labeling) listening modes. However, contrary to their hypothesis and the results from Pisoni (1973), the authors do not find that longer inter-stimulus intervals in the 2I2AFC task yield increasingly categorical perception.

The task demand manipulations that yield more continuous or categorical discrimination patterns in the studies reviewed above are hypothesized to be associated with acoustic and phonological listening modes, i.e., processing limited to acoustic differences and processing that also integrates phonological knowledge, respectively. The phonological listening mode involves phonological representations, such as the phonological contrast between two speech sounds by which they are phonemically labeled as two different speech categories, and integrates top-down influences on speech processing such as lexical information. A phoneme identification study by Fox (1984) presents evidence showing that lexical influences on speech perception become evident later than acoustic influences of the speech signal, suggesting distinct mechanisms for acoustic and phonological processing. In Fox's (1984) phoneme identification study, listeners were presented with steps along a word-nonword continuum such as /bæg-dæg/ or /bæb-dæb/, where /bæg/ and /dæb/ are words and /dæg/ and /bæb/ are nonwords, respectively, and were asked to identify the initial consonant in each stimulus. Responses were grouped by response time ranges (responses faster than 500 ms, responses between 500 and 800 ms, and responses slower than 800 ms). A greater proportion of slower responses were ‘word’ responses in each continuum compared to faster responses, suggesting that the top-down lexicality effects (Ganong, 1980) are less prominent in the earlier stages of phonemic categorization. This finding converges with Pisoni (1973) that longer processing of speech sounds involve a greater degree of phonological processing.

Similarly, Johnson and Babel (2010) and Babel and Johnson (2010) show that in 'speeded' AX consonant discrimination tasks with 100 ms inter-stimulus intervals and instructions and feedback at each trial to encourage responses under 500 ms, whether the consonants to be discriminated were phonemically contrastive or allophonic in the native languages of the listeners did not influence response times. Moreover, response times in 'speeded' AX consonant discrimination were correlated with participants' similarity ratings of the same stimulus pairs with 100 ms inter-stimulus intervals, suggesting that under task demands to make fast discrimination responses, the variation in response times as a measure of perceptual distance reflected acoustic similarity of the stimuli that is not language specific. In contrast, when listeners were slower in their responses, both similarity ratings and response times in discrimination reflected language-specific phonemic status (Boomershine et al., 2008).

The findings reviewed above confirm that both vowels and consonants can be perceived continuously or categorically, reflecting acoustic and phonological processing of speech, respectively. The task demand manipulations that target these speech perception processes can be leveraged to ask questions about and gain a better understanding of speech perception. In this vein, Johnson and colleagues present a series of perceptual discrimination studies with varying task demands, aiming to reveal the relative contributions of acoustic-phonetic factors and phoneme inventory- and feature-based phonological factors to speech perception. In Ettlinger and Johnson (2010), native English, French, and Turkish listeners performed two AX vowel discrimination tasks and a similarity rating task on German vowels /i ɪ y ʏ/. The researchers were interested in the influence of native vowel inventories and native vowel feature contrasts on nonnative vowel discrimination. Of these four German vowels, /i/ is native to all three test languages, /y/ is native to French and Turkish, /ɪ/ is native to English, and /ʏ/ is not native to any of the test languages. On the other hand, whereas French and Turkish have a rounding contrast within front vowels, English has a tenseness contrast within front vowels. If listeners engage in feature-based processing, it is predicted that they should be faster in discriminating vowels that have a feature contrast that they are familiar with even when the vowels are nonnative and hence not phonemically identifiable.

Listeners performed a speeded AX discrimination task with a short inter-stimulus interval (100 ms) and received feedback based on the acoustic rather than phonemic identity of the stimuli, aiming to encourage acoustic rather than phonological processing of the stimuli and to obtain a measure of perceived vowel similarity independent of linguistic experience. The three language groups did not behave differently in the speeded discrimination task, suggesting that the speeded AX discrimination task responses reflected the acoustic similarity of the vowel tokens. Listeners also performed an AX vowel discrimination task with long inter-stimulus intervals (500 ms) and received feedback based on the phonemic identity of the stimuli, aiming to encourage responses based on phoneme identification and to obtain a measure of perceived vowel similarity reflecting native language phonology. In this task, Turkish and French listeners were significantly slower than English listeners when discriminating the vowel pair /ɪ ʏ/, suggesting that Turkish and French listeners' inability to phonemically identify either of these vowels made these vowels perceptually more similar and less discriminable, although the vowel pair /y ʏ/ did not yield slower responses from English listeners compared to Turkish and French listeners. The authors argue that the perceived similarity of nonnative vowel pairs in this task was largely predicted by their acoustic similarity, with the phonetically distant vowel pair /y ʏ/ receiving fast responses from all language groups. In contrast, the phonetically close vowel pair /ɪ ʏ/ showed an additional influence of native vowel phonology. In conclusion, contrary to what was hypothesized, long inter-stimulus intervals and feedback based on phonemic identity in the AX vowel discrimination task with long intervals did not consistently target phonological processing of nonnative vowels for all language groups.

Relatedly, Johnson (2015) utilized varying task demands in AX vowel discrimination tasks and presented native English listeners with German vowels /e o ø/ to explore the influence of native vowel phonology on vowel perception in comparison to acoustic similarity. In the 'auditory' discrimination task, the inter-stimulus intervals were 100 ms long and feedback was based on the acoustic identities of the stimuli. In the 'phonemic' discrimination task, the interstimulus intervals were 700 ms long and feedback was based on the phonemic identities of the stimuli. Response times in the 'phonemic' discrimination task were longer than those in the 'auditory' discrimination task, suggesting a difference in vowel processing across the two tasks. Moreover, these differences in response times were not equal across the vowel pairs and were predicted by the unique features of the vowels in each pair, suggesting that the task demands of the 'phonemic' discrimination task engaged a feature-based phonological processing of the nonnative vowels.

The mixed results from Ettlinger and Johnson (2010) suggest that even with longer inter-stimulus intervals, response times in AX vowel discrimination tasks might reflect the acoustic properties of the stimuli and not phonological processing. An alternative approach when aiming to compare acoustic and phonological processing might be to not only manipulate the inter-stimulus intervals and the type of feedback within a task but to compare performance across tasks that rely on detecting acoustic differences regardless of phoneme labels and tasks that require phoneme labeling, such as in Gerrits and Schouten's (2004) comparison between oddball recognition in 4I2AFC and phoneme label order discrimination in 2I2AFC tasks. In a study examining whether the perception of nonnative consonant clusters involve phonetic or phonological processing, Davidson and Shaw (2012) compared discrimination performance in AX discrimination and ABX discrimination tasks, in line with the work reviewed above suggesting that AX discrimination tasks more readily target acoustic processing and ABX discrimination tasks more readily target phonological processing (e.g., Pisoni, 1973). Their results suggested that AX discrimination patterns reflected continuous perception and were explained by the acoustic similarity of the stimuli, whereas ABX discrimination patterns reflected language-specific phonological and phonotactic knowledge. Thus, Davidson and Shaw's (2012) results suggest that combining AX and ABX discrimination tasks might enable a comparison of acoustic and phonological processing of speech, which was not achieved in Ettlinger and Johnson's (2010) comparison with AX vowel discrimination tasks. In the forthcoming, following Davidson and Shaw (2012), I present a cross-linguistic comparison of Turkish and English AX and ABX vowel discrimination to ask questions about the phonological and phonotactic predictors of perceived vowel similarity in each language beyond acoustic predictors of perceived similarity.

3.2. The present study

Based on the studies reviewed above, to examine the types of information that are integrated in the phonological processing of vowels beyond the vowel acoustics, I compare response times in native AX and ABX vowel discrimination tasks by English and Turkish listeners, where I manipulate task type (AX or ABX vowel discrimination), inter-stimulus interval (short or long), feedback type (acoustic or phonemic), and speaker identity in vowel discrimination (same speaker or different speakers) to target acoustic or phonological listening modes. I hypothesize that the response times in the AX vowel discrimination task with short inter-stimulus intervals, feedback based on acoustic identity and stimuli produced by a single speaker will parallel the measures of acoustic similarity of the stimuli, whereas response times in the ABX vowel discrimination task with long inter-stimulus intervals, feedback based on phonemic identity and stimuli produced by two different speakers will be influenced by phonological and phonotactic knowledge on top of acoustic similarity.

By comparing response times in AX and ABX vowel discrimination tasks and modeling response times in the ABX vowel discrimination task with both acoustic and phonological predictors, I aim to examine whether vowel feature-based similarity and phonotactic similarity (defined in terms of the likelihood of vowel cooccurrence in the lexicon) influence perceived vowel similarity in Turkish beyond the acoustic similarity of the vowels. For the effect of likelihood of vowel cooccurrence in the Turkish lexicon on perceived vowel similarity, I hypothesize that vowels that are more likely to cooccur will be coactivated to a higher degree and hence receive slower responses when presented as a pair in a discrimination task trial. In the cross-linguistic comparison of Turkish vowel discrimination with English vowel discrimination, I hypothesize that the response times in ABX vowel discrimination tasks will reflect a greater influence of phonological predictors, including the likelihood of vowel cooccurrence, compared to AX vowel discrimination in both Turkish and English. On the other hand, I hypothesize that the likelihood of vowel cooccurrence in their respective lexicons will be more influential in Turkish vowel discrimination than in English vowel discrimination.

3.3. Turkish vowel discrimination

3.3.1. Methods

3.3.1.1. Participants

Participants were 40 monolingually raised native speakers of Turkish aged 18-61 (M_{age} = 26.35; 24 women, 1 other), residing in Istanbul, Turkey, at the time of participation. Thirty-five participants self-reported English as a second language, and 32 of these participants self-reported they were fluent in English. One participant self-reported German as a second language they were fluent in. Participants' self-reported age of onset for English was ranged from 5 to 21 (M_{age} = 9.47). Sixteen participants self-reported knowing a third language, 7 participants self-reported knowing a fourth language, and 1 participant self-reported knowing a fifth language. More specifically, 6 participants knew Arabic, 5 participants knew French, 4 participants knew German, 3 participants knew Korean, 3 participants knew Japanese, 1 participant knew Spanish, 1 participant knew Chinese, and 1 participant knew Indonesian. Participants were paid ₺300 Turkish Liras for their participation at the completion of the experiment.

3.3.1.2. Stimuli

3.3.1.2.1. AX vowel discrimination

Stimuli for the AX vowel discrimination task consisted of 24 monosyllabic /bVb/ nonword tokens with 3 tokens per each of the 8 Turkish vowels (/ɑ e ɯ i o œ u y/) produced by one of the female speakers in the Turkish vowel production study reported in Chapter 2. The 3 tokens per vowel were selected such that they would have varying and non-overlapping vowel formants to facilitate the discrimination of the different tokens of a vowel (Figure 3.1). The vowel tokens of interest were extracted from the audio recordings and the vowel duration, intensity, and pitch contours were manipulated using Praat to control these variables across the stimuli. All vowel stimuli were manipulated to be 97 ms long (mean duration of the vowels in the monosyllabic nonwords in the stimuli) with a mean intensity of 70 dB. For each vowel token, the naturally produced pitch was replaced with a falling pitch contour extracted from a natural production of one of the stimulus tokens. Vowel formants were not manipulated, to ensure naturalness of the stimuli.

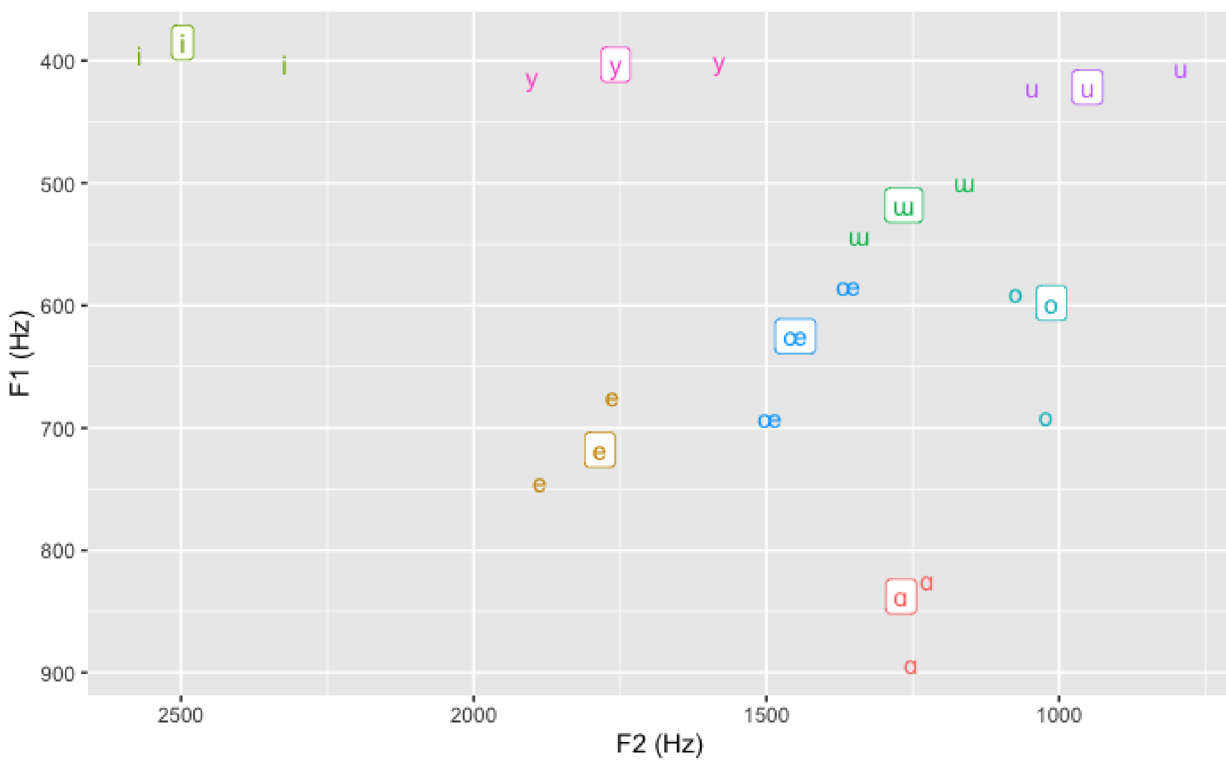

Figure 3.1. Vowel formants of the vowel stimuli for AX discrimination experiment, measured at vowel midpoint. All three tokens of each vowel occurred in the 'same vowel' condition. The vowel tokens depicted in boxes were paired with each other in the 'different vowel' condition.

The stimulus tokens were concatenated in pairs with a 100 ms silent inter-stimulus interval and a 500 ms silent buffer before the first stimulus token. There were three sets of AX vowel discrimination stimulus pairs: same token stimulus pairs, same vowel different token stimulus pairs, and different vowel stimulus pairs. There were 24 same token stimulus pairs where A and X are acoustically identical stimuli. All 3 tokens of the 8 vowels were combined with themselves in these stimulus pairs. There were 48 same vowel different token stimulus pairs where A and X are phonemically identical but acoustically distinct stimuli. For each of the 8 vowels, the 3 tokens of the vowel were presented in the possible permutations (P(3, 2)), yielding 6 stimulus pairs per vowel. Lastly, there were 56 different vowel stimulus pairs where A and X are phonemically (and hence also acoustically) distinct stimuli. For each of the 8 vowels, one token with the least ambiguous vowel formants per vowel was selected and the tokens were presented in the possible permutations (P(8, 2)), yielding 7 stimulus pairs per vowel where the vowel is stimulus A and 7 stimulus pairs per vowel where the vowel is stimulus X.

3.3.1.2.2. ABX vowel discrimination

Stimuli for the ABX vowel discrimination task consisted of 16 monosyllabic /bVb/ nonword tokens with 1 token per each of the 8 Turkish vowels (/ɑ e ɯ i o œ u y/) produced by two female speakers in the production experiment reported in Chapter 2. The vowel tokens from Speaker 1 were used only in the ‘A’ and ‘B’ stimuli in the stimulus triplets and the vowel tokens from Speaker 2 were used only in the ‘X’ stimuli in the stimulus triplets (Figure 3.2). The vowel tokens of interest were extracted from the audio recordings and the vowel duration, intensity, and pitch contours were manipulated using Praat to control these variables across the stimuli. All vowel stimuli were manipulated to be 97 ms long (mean duration of the vowels in the monosyllabic nonwords in the stimuli) with a mean intensity of 70 dB. The pitch contour of all vowel tokens was replaced with a falling pitch tier that was extracted from one of the naturally produced stimulus tokens. Vowel formants were not manipulated, to ensure naturalness of the stimuli.

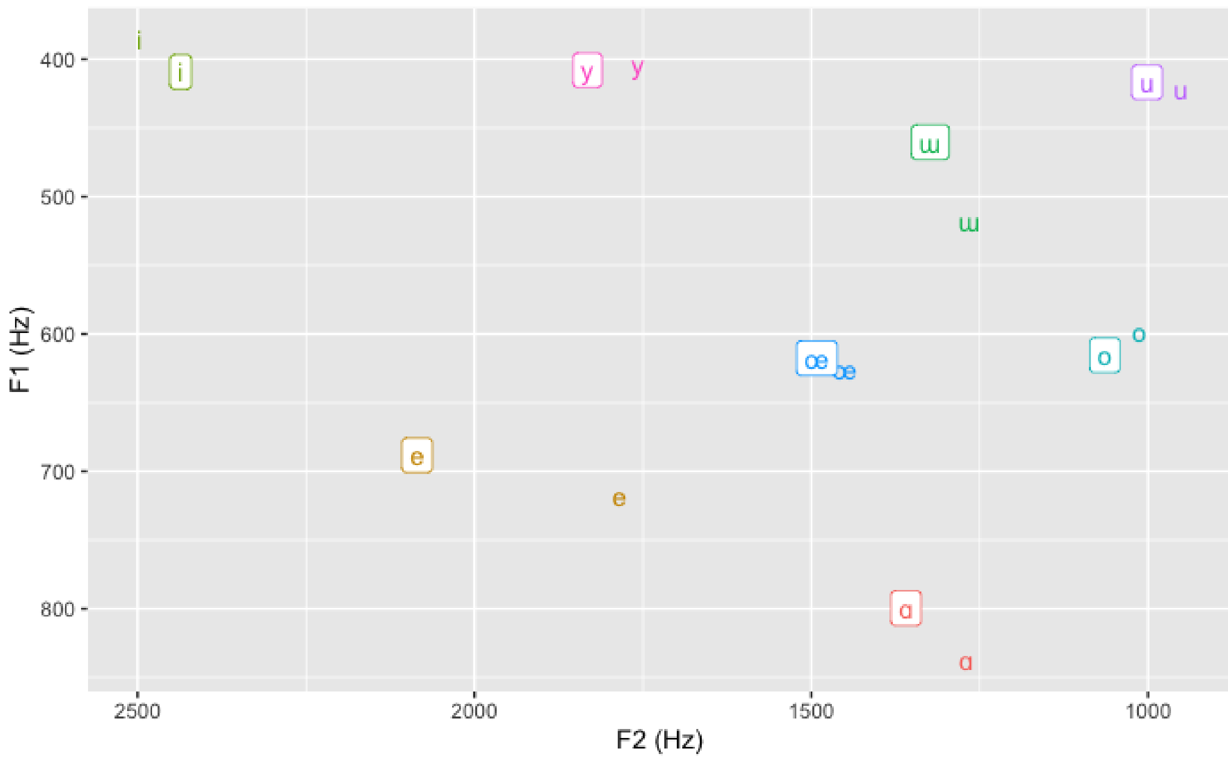

Figure 3.2. Vowel formants of the vowel stimuli for ABX discrimination experiment, measured at vowel midpoint. The vowel tokens depicted in boxes are the vowel tokens that were produced by Speaker 1 and that occurred in the A and B positions of the stimulus triplet, and the vowel tokens depicted without boxes are the vowel tokens that were produced by Speaker 2 and that occurred in the X position of the stimulus triplet.

The stimulus tokens were concatenated in triplets with a 500 ms silent inter-stimulus interval and a 500 ms silent buffer before the first stimulus token. There were two types of ABX vowel discrimination stimulus triplets: A = X triplets and B = X triplets. There were 56 triplets in which the vowel in the target syllable ‘X’ was the same vowel phoneme as the vowel in the first syllable ‘A’, and 56 triplets in which the vowel in the target syllable ‘X’ was identical with the vowel in the second syllable ‘B’.

3.3.1.3. Procedure

The vowel perception study was conducted in person at reserved rooms in school campuses in Istanbul, Turkey. Participants were seated in front of a MacBook Air laptop computer with a touchpad and were instructed to wear corded headphones. Participants first gave their informed consent and completed a language background survey, after which they completed the vowel perception study consisting of a vowel identification task (see Chapter 4 for details), the AX vowel discrimination task and lastly the ABX vowel discrimination task. Participants completed the three vowel perception experiments in a single session and could take breaks during the session with optional break screens halfway through each experiment.

3.3.1.3.1. AX discrimination task

There were 208 experimental trials in the AX discrimination task and participants completed 4 practice trials before beginning the experimental trials. Each trial began with a fixation cross visually presented on screen for 1 second. In each trial, participants were auditorily presented with a randomly selected stimulus pair. All stimulus pairs consisted of a 500 ms silent buffer, the auditory presentation of the first stimulus (A) followed by the auditory presentation of the second stimulus (X) with a 100 ms silent inter-stimulus interval between A and X. Immediately after the auditory presentation of the stimulus pair, a response screen was presented visually, asking participants to indicate whether the two audio recordings they were presented with were identical or different. The two responses were assigned the ‘F’ and ‘J’ keys on the keyboard and the response keys were reminded to the participant on the response screen on each trial. This response-key assignment was counter-balanced across participants.

Participants were instructed that in each trial, they would hear two audio recordings with a one-syllable made-up word spoken by the same speaker played back-to-back, and that they would be asked to indicate whether they thought the two audio recordings were identical or different by pressing ‘F’ and ‘J’ keys on the keyboard as was shown on the response screen. Participants were instructed to listen for subtle differences in the speaker’s pronunciation of the made-up words for their responses. It was explained to participants that there are a number of cases they might encounter: If they hear different sounds such as ‘beb’ followed by ‘bab’, the two audio recordings are ‘different’ because the sounds are different. If they hear ‘beb’ followed by a second ‘beb’ and although it is the same made-up word, the speaker said the word a little differently each time in the two audio recordings, the two audio recordings are nonetheless considered ‘different’. Lastly, if they hear ‘beb’ twice in a row and if both audio recordings sound identical in every detail, they are considered ‘identical’ audio recordings. Participants were asked to respond as quickly and as accurately as possible and pick the response that seems best if they are not sure.

Upon participant response, their response was highlighted on screen, and they received feedback on whether the two audio recordings in that trial were identical or different. Participants were told that the feedback is not intended to be about their performance or ‘accuracy’ as the only right answer is the one that reflects their perception, but that the feedback serves as a reminder about listening for subtle differences as two audio recordings with the same word might nonetheless be ‘different’. The feedback remained on screen for 2 seconds, after which a new trial began automatically. Participants were instructed to keep their left hand on the ‘F’ key and their right hand on the ‘J’ key throughout the experiment as trials advanced and stimuli played automatically.

3.3.1.3.2. ABX discrimination task

There were 112 experimental trials in the ABX discrimination task and participants completed 4 practice trials before beginning the experimental trials. Each trial began with a fixation cross visually presented on screen for 1 second. In each trial, participants were auditorily presented with a randomly selected stimulus triplet. All stimulus triplets consisted of a 500 ms silent buffer and the auditory presentation of the three stimuli (A, B, and X) with 500 ms silent inter-stimulus intervals between each of the A, B, and X stimuli. Immediately after the auditory presentation of the stimulus triplet, a response screen was presented visually, asking participants to indicate whether the vowel they heard in the third and last syllable was identical with the vowel in the first syllable or the second syllable. The 'first syllable' response was assigned to the 'F' key and the 'second syllable' response was assigned to the 'J' key on the keyboard, and the response keys were reminded to the participant on the response screen in each trial. Participants were instructed that the first two syllables were spoken by the same speaker and the third syllable was spoken by a different speaker. Participants were asked to respond as quickly and as accurately as possible and pick the response that seems best if they are not sure.

Upon participant response, their response was highlighted on screen, and they received feedback on whether the vowel in the third syllable was identical with the vowel in the first syllable or the second syllable. Participants were told that the feedback is not intended to be about their performance or 'accuracy' as the only right answer is the one that reflects their perception, but that the feedback serves as a reminder about listening for identical vowels in the tree syllables. The feedback remained on screen for 2 seconds, after which a new trial began automatically. Participants were instructed to keep their left hand on the 'F' key and their right hand on the 'J' key throughout the experiment as trials advanced and stimuli played automatically.

3.3.2. Results

3.3.2.1. AX discrimination task

In the AX discrimination task, critical trials are the trials in which the stimulus pair consists of two phonemically distinct vowels. There were 4480 critical trials and participants responded accurately to 98.64% of the critical trials. Trials in which an inaccurate response was made and trials in which response times were more than 2.5 standard deviations above the mean response time were excluded (M_{RT} = 866.09 ms, SD_{RT} = 575.19 ms). Response times (log-transformed) from the second stimulus offset for a total of 4404 accurately responded trials were modeled in a linear mixed effects regression (13).

(13) response time ~ response key
+ Euclidean distance
+ Pillai score
+ feature dissimilarity ratio
+ PMI
+ (1 + Euclidean distance + Pillai score + feature dissimilarity ratio
+ PMI | participant)

Predictors in the model were response key (whether the accurate 'different' response was made with the 'F' key and the left hand or the 'J' key and the right hand), Euclidean distance between stimulus vowel pair, Pillai score as a measure of overlap between the stimulus pair vowel categories in F1-F2 space (calculated via a MANOVA over the full monosyllabic nonword vowel productions described in Chapter 2), vowel feature dissimilarity ratio between the stimulus pair vowel categories (calculated as the number of features that the two vowels differ in divided by 3, the total number of Turkish vowel features), and PMI between the stimulus pair vowel categories (see Chapter 1). Following a stepwise model selection process (Barr et al., 2013), the mixed effects models in this work had maximal fixed and random effects where possible, and random slopes or intercepts were dropped when the maximal model failed to converge. In this model, there were random intercepts by participant and random slopes by Pillai score and PMI. No other random intercepts or slopes were included in the model due to singularity and/or convergence issues. Categorical predictors were sum coded and continuous predictors were centered and standardized.

There was a significant main effect of Pillai score ($\beta = -0.02$, $SE = 0.01$, $t = -2.27$, $p = 0.03$), indicating that participant responses became faster with increasing Pillai score indexing lesser overlap between the stimulus pair vowel categories (Figure 3.3). There was also a significant main effect of feature dissimilarity ratio ($\beta = -0.02$, $SE = 0.01$, $t = -2.17$, $p = 0.03$), indicating that participant responses became faster with increasing dissimilarity in the vowel features of the stimulus vowel categories (Figure 3.4). None of the other predictors had significant effects on response times, with the effect of PMI approaching significance ($\beta = -0.01$, $SE = 0.01$, $t = -1.90$, $p = 0.06$), however, this effect was not in the predicted direction with faster participant responses for stimulus vowel categories with greater PMI or greater likelihood of cooccurrence (Table 3.1; see Chapter 4 for a discussion of a similar effect). Overall, these results suggest that Turkish listeners engage in processing that involves vowel categories and features in the AX vowel discrimination task. Moreover, contrary to my hypothesis about the speeded AX vowel discrimination task targeting acoustic-phonetic processing, I do not find evidence that the phonetic distance measure predicts response times in this task. I argue that any potential influence of Euclidean distance might be masked by Pillai scores, as both are vowel similarity measures based on phonetic distance and as the two predictors have the highest correlation among the fixed effects ($R = -0.59$). Nonetheless, that Pillai score was a stronger predictor of response times than feature dissimilarity ratio suggests that participants' processing of the AX discrimination task was not purely phonological and involved phonetic dissimilarity despite both predictors referencing phonemic vowel categories rather than individual vowel tokens.

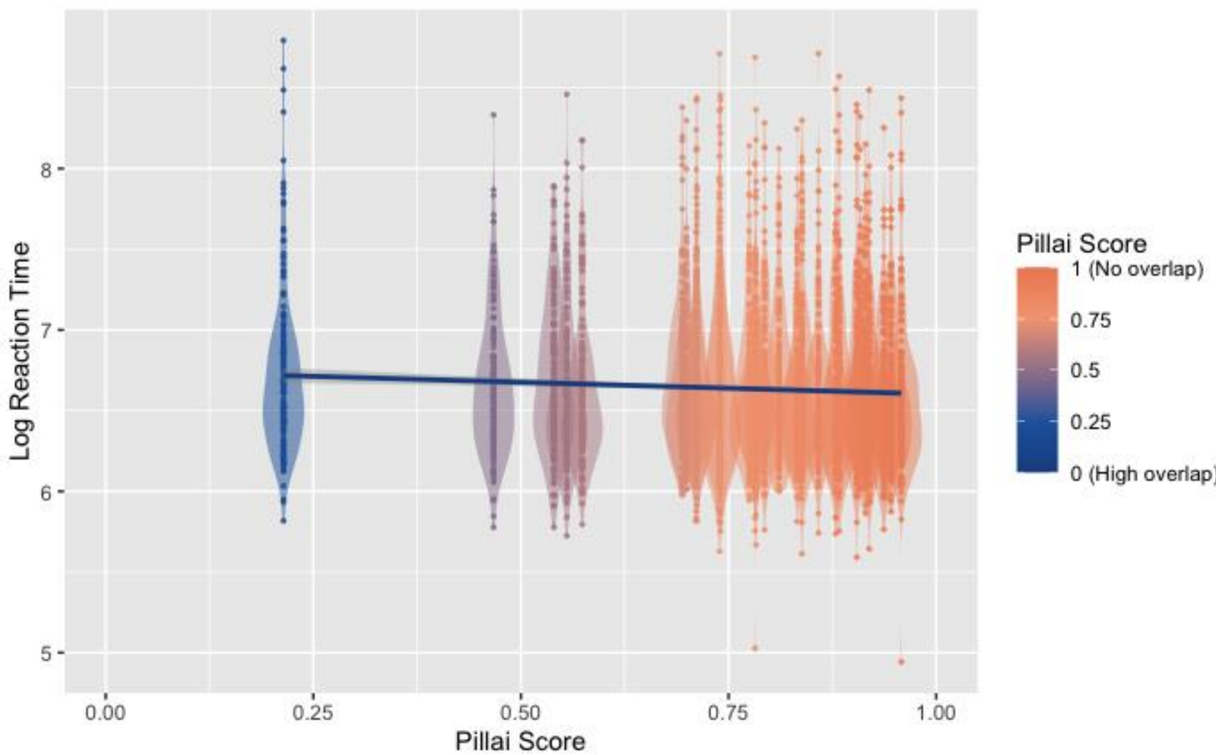

Figure 3.3. Distribution of response times (log transformed) as a function of Pillai scores of stimulus pairs in the AX vowel discrimination task, with each point representing one trial. Fitted line represents the linear relationship between response time and Pillai score.

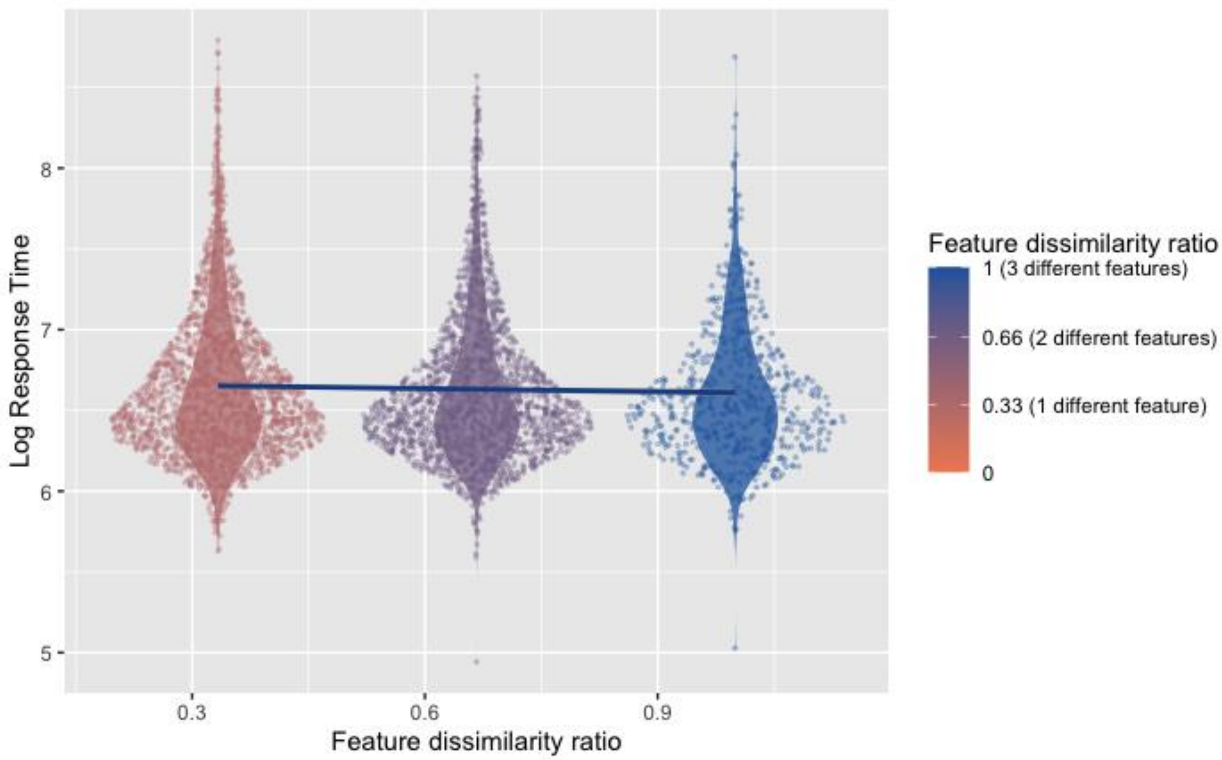

Figure 3.4. Distribution of response times (log transformed) as a function of the vowel feature dissimilarity ratio of stimulus pairs in the AX vowel discrimination task, with each point representing one trial. Fitted line represents the linear relationship between response time and the vowel feature dissimilarity ratio.

Table 3.1. Model results for AX vowel discrimination task.

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	6.65	0.04	39.74	165.96	< 0.0001***
Response key	-0.03	0.04	39.59	-0.69	0.49
Euclidean distance	0.001	0.01	41.07	0.11	0.92
Pillai score	-0.02	0.01	41.25	-2.27	0.03*
Feature dissimilarity ratio	-0.02	0.01	98.94	-2.17	0.03*
PMI	-0.01	0.01	39.70	-1.90	0.064

3.3.2.2. ABX discrimination task

In the ABX discrimination task, there was a total of 4480 trials and participants responded accurately to 94.87% of the trials. Trials in which an inaccurate response was made and trials in which response times were more than 2.5 standard deviations above the mean response time were excluded (M_{RT} = 1138.03 ms, SD_{RT} = 712.15 ms). Response times (log-transformed) from the third stimulus offset for a total of 4068 accurately responded trials were modeled in a linear mixed effects regression (14).

(14) response time ~ stimulus order
+ Euclidean distance
+ Pillai score
+vowel feature dissimilarity ratio
+ PMI
+ (1 + stimulus order + Pillai score | participant)
+ (1 | item)

Predictors in the model were stimulus order (whether the vowel in the target syllable 'X' was identical with the vowel in the first syllable 'A' or the vowel in the second syllable 'B'), Euclidean distance between the vowels in 'A' and 'B' in the F1-F2 space, Pillai score as a measure of overlap between the vowel categories in 'A' and 'B' in the F1-F2 space, feature dissimilarity ratio between the vowel categories in 'A' and 'B' (calculated as the number of features that the two vowels differ in divided by 3, the total number of Turkish vowel features), and PMI between the vowel categories in 'A' and 'B'. There were random intercepts by participant with random slopes by stimulus order and Pillai score and random intercepts by item. No other random slopes were included in the model due to singularity issues. Categorical predictors were sum coded and continuous predictors were centered and standardized.

There was a significant main effect of stimulus order (β = -0.04, SE = 0.01, t = -2.45, p = 0.02), indicating that participant responses were faster when the vowel in the target syllable 'X' was identical with the vowel in the second syllable 'B' compared to the first syllable 'A'. None of the other predictors had significant effects on response times (Table 3.2). This result suggests that hearing syllables with identical vowels consecutively in the second and third stimuli of the stimulus triplet had an advantage in processing. As this processing difference between the vowels in the first syllable 'A' and the second syllable 'B' may have obscured the effects of the other predictors, I also modeled the results separately by stimulus order.

Table 3.2. Model results for ABX discrimination task.

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	6.90	0.05	39.76	153.21	< 0.0001***
Stimulus order	-0.04	0.01	46.09	-2.45	0.02*
Euclidean distance	-0.004	0.01	109.52	-0.30	0.76
Pillai score	-0.02	0.01	82.01	-1.46	0.15
Feature dissimilarity ratio	-0.01	0.009	110.78	-1.12	0.26
PMI	0.002	0.009	110.72	0.19	0.85

For the A = X trials in which the vowel in the target syllable 'X' was identical with the vowel in the first syllable 'A', the predictors in the model (15) were Euclidean distance between the vowels in 'A' and 'B' in the F1-F2 space, Pillai score as a measure of overlap between the vowel categories in 'A' and 'B' in the F1-F2 space, feature dissimilarity ratio between the vowel categories in 'A' and 'B' (calculated as the number of features that the two vowels differ in divided by 3, the total number of vowel features), and PMI between the vowel categories in 'A' and 'B'. There were random intercepts by participant and item, and random slopes for Pillai score by participant. No other random slopes were included in the model due to singularity and convergence issues.

(15) response time ~ Euclidean distance

+ Pillai score

+vowel feature dissimilarity ratio

+ PMI

+ (1 + Pillai score | participant)

+ (1 | item)

There was a near significant main effect of feature dissimilarity ratio ($\beta = -0.02$, $SE = 0.01$, $t = -1.95$, $p = 0.056$), indicating that participant responses became faster with increasing dissimilarity in vowel features of the first syllable 'A' and second syllable 'B' vowel categories (Figure 3.5). In other words, discrimination between the vowel in the first syllable 'A' and the second syllable 'B' was facilitated when these two vowels had fewer vowel features in common. None of the other predictors had significant effects on response times (Table 3.3).

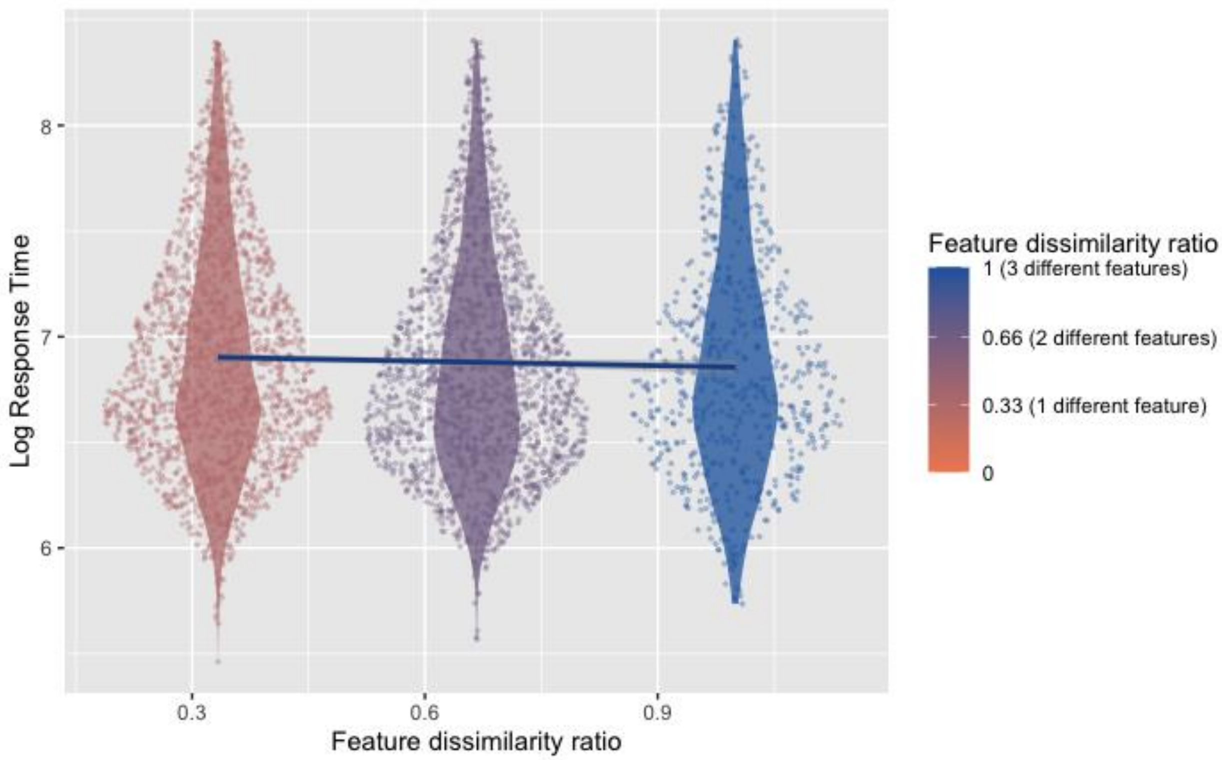

Figure 3.5. Distribution of response times (log transformed) as a function of A-B vowel feature dissimilarity ratio of A = X stimulus triplets in the ABX vowel discrimination task, with each point representing one trial. Fitted line represents the linear relationship between response time and vowel feature dissimilarity ratio between A and B vowels.

Table 3.3. Model results for A = X trials in the ABX discrimination task.

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	6.93	0.05	39.06	137.10	< 0.0001***
Euclidean distance	0.01	0.02	54.18	0.85	0.39
Pillai score	-0.03	0.02	43.08	-1.59	0.11
Feature dissimilarity ratio	-0.02	0.01	55.33	-1.95	0.056
PMI	-0.004	0.01	55.95	-0.35	0.73

For the B = X trials in which the vowel in the target syllable ‘X’ was identical with the vowel in the second syllable ‘B’, the predictors in the model (16) were Euclidean distance between the vowels in ‘A’ and ‘B’ in the F1-F2 space, Pillai score as a measure of overlap between the vowel categories in ‘A’ and ‘B’ in the F1-F2 space, feature dissimilarity ratio between the vowel categories in ‘A’ and ‘B’ (calculated as the number of features that the two vowels differ in divided by 3, the total number of vowel features), and PMI between the vowel categories in 'A' and 'B'. There were random intercepts by participant and item, and random slopes for feature dissimilarity ratio by participant and Pillai score by item. No other random slopes were included in the model due to singularity and convergence issues.

(16) response time ~ Euclidean distance
+ Pillai score
+ feature dissimilarity ratio
+ PMI
+ (1 + feature dissimilarity ratio | participant)
+ (1 + Pillai score | item)

None of the predictors were found to have significant effects on response times (Table 3.4).

Table 3.4. Model results for B = X trials in the ABX discrimination task.

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	6.86	0.04	41.51	158.41	< 0.0001***
Euclidean distance	-0.02	0.02	48.81	-1.01	0.32
Pillai score	-0.01	0.02	15.66	-0.92	0.37
Feature dissimilarity ratio	0.005	0.01	41.71	0.36	0.71
PMI	0.007	0.01	53.04	0.59	0.56

In addition to the faster responses in the B = X trials compared to the A = X trials that was found in the aggregate model, accuracy rates also varied between the two types of trials. I modeled accuracy in the ABX vowel discrimination task in a binomial logistic regression with stimulus order, Euclidean distance, Pillai score, vowel feature dissimilarity ratio, and PMI as predictors. Stimulus order was the only significant predictor of accuracy in this model ($\beta = 0.26$, $SE = 0.07$, $z = 3.67$, $p = 0.0002$), indicating that B = X trials received more accurate responses compared to A = X trials, suggesting a response bias favoring 'second syllable' responses. Hence, I argue that the two trial types are processed differently by participants, with the A = X trials being harder compared to the B = X trials and greater feature dissimilarity in A and B vowels facilitating vowel discrimination in the harder A = X trial type. Overall, these results suggest that Turkish listeners possibly engage in processing that involves vowel categories and features in the ABX vowel discrimination task.

3.3.3. Discussion

This vowel discrimination study aims to compare response times in two vowel discrimination tasks with task demands that are hypothesized to differ in the relative importance of phonetic as opposed to phonological processing. Under the assumption that the ABX vowel discrimination task involves relatively more phonological processing than AX vowel discrimination task, the main questions of interest are whether Turkish listeners utilize vowel features (in other words, harmony classes) in phonological processing of Turkish vowels and whether lexical vowel cooccurrence statistics predicts coactivation of Turkish vowel categories, with vowels with higher likelihood of cooccurrence being coactivated and perceived to be more similar, in effect being harder to discriminate.

The results comparing AX and ABX vowel discrimination tasks partially confirm the hypothesis about the listening modes or levels of processing involved in the two tasks: Pillai scores were a significant predictor of response times in the AX vowel discrimination task, indicating that listeners processed the vowel tokens with respect to the phonetic overlap of their respective vowel categories, whereas vowel feature dissimilarity ratio, a phonological predictor, was a significant predictor of response times in the AX vowel discrimination task and a marginally significant predictor of response times in the A = X trials in the ABX vowel discrimination task. Although the AX vowel discrimination task was designed to not involve phonemic vowel identification, the results suggest that listeners engaged in phonological vowel processing in this task. Nevertheless, that response times in this task reflect phonological predictors suggests that perceived vowel similarity cannot be explained by phonetic predictors alone and that phonological factors contribute to perceived vowel similarity beyond phonetic factors.

That the only significant predictor in the ABX vowel discrimination task was stimulus order suggests that the two trial types reflect two different processing strategies. Listeners may have paid attention to and compared only the B and X stimuli, responding with ‘second syllable’ when they match and only considering stimulus A and respond with ‘first syllable’ when they mismatch, regardless of how they would label the vowels. Although I have not asked participants to report their strategies, one participant communicated upon completing the experiment that they only paid attention to the B and X stimuli and disregarded stimulus A in the ABX vowel discrimination task.

Recall that we are interested in the results from the ABX vowel discrimination task for evidence that harmony-related phonological predictors play a role in vowel perception. However, greater vowel feature dissimilarity ratio was a significant predictor of faster responses in the AX vowel discrimination task. The fact that vowel feature dissimilarity ratio is a significant predictor of response times supports my hypothesis that Turkish listeners rely on phonological vowel features, which are relevant in Turkish vowel harmony. However, PMI was not a significant predictor in either vowel discrimination task, suggesting that likelihood of vowel cooccurrence in the Turkish lexicon does not influence perceived vowel similarity for Turkish listeners when discriminating vowels. That is, whereas the significant effect of vowel feature dissimilarity ratio suggests that Turkish listeners invoke vowel feature representations in vowel discrimination, I do not find evidence that they also invoke vowel cooccurrence relationships that I hypothesized to underlie the vowel feature representations.

To reiterate, I hypothesize that Turkish vowel harmony enhances perceived vowel similarity within a harmony class (also see Terbeek, 1977), thereby contributing to the salience of vowel features in Turkish vowel perception. In contrast, in a nonharmony language, vowels do not form harmony classes that can be defined by vowel features. Hence, I hypothesize that the significant relationship between vowel feature dissimilarity ratio and perceived vowel similarity to be stronger in Turkish than in a nonharmony language. To test this hypothesis, I provide a cross-linguistic comparison with English vowel discrimination in the following.

3.4. English vowel discrimination

3.4.1. Methods

3.4.1.1. Participants

Participants were 40 monolingually raised native speakers of English aged 18-63 (M_{age} = 31.50; 22 women). Fourteen participants self-reported knowing a second language, and 2 participants self-reported knowing a third language. More specifically, 7 participants knew Spanish, 6 participants knew French, 1 participant knew German, 1 participant knew Indonesian, and 1 participant knew Portuguese. Earliest self-reported age of onset for a second or third language was 6. Twenty-seven participants were recruited from Prolific, an online recruitment platform, and were paid $10 USD for their participation upon completing of the experiment. Twenty-three participants were recruited from the Linguistics Subject Pool at Northwestern

University and were given course credit for their participation. Data from 20 other participants were excluded due to reporting a native language other than English or bilingual development.

3.4.1.2. Stimuli

3.4.1.2.1. AX vowel discrimination

Stimuli consisted of 24 monosyllabic /bVp/ tokens with 3 tokens per each of the 8 English vowels (/ɑ e ɛ i ɪ o u ʊ/) produced by one female native English speaker. The 3 tokens per vowel were selected such that they would have varying and non-overlapping vowel formants to facilitate the discrimination of the different tokens of a vowel. The vowel tokens of interest were extracted from the audio recordings and the vowel duration, intensity, and pitch contours were manipulated using Praat to control these variables across the stimuli. All lax vowel stimuli (/ɛ ɪ ʊ/) were manipulated to be 100 ms long (mean duration of lax vowels produced by the speaker) and all tense vowel stimuli (/ɑ e i o u/) were manipulated to be 130 ms long (mean duration of the tense vowels produced by the speaker). All vowel stimuli's mean intensity was set to 70 dB and pitch tiers were replaced with a naturally falling pitch tier that was extracted from one of the stimulus tokens. Vowel formants were not manipulated to ensure naturalness of the stimuli (Figure 3.6).

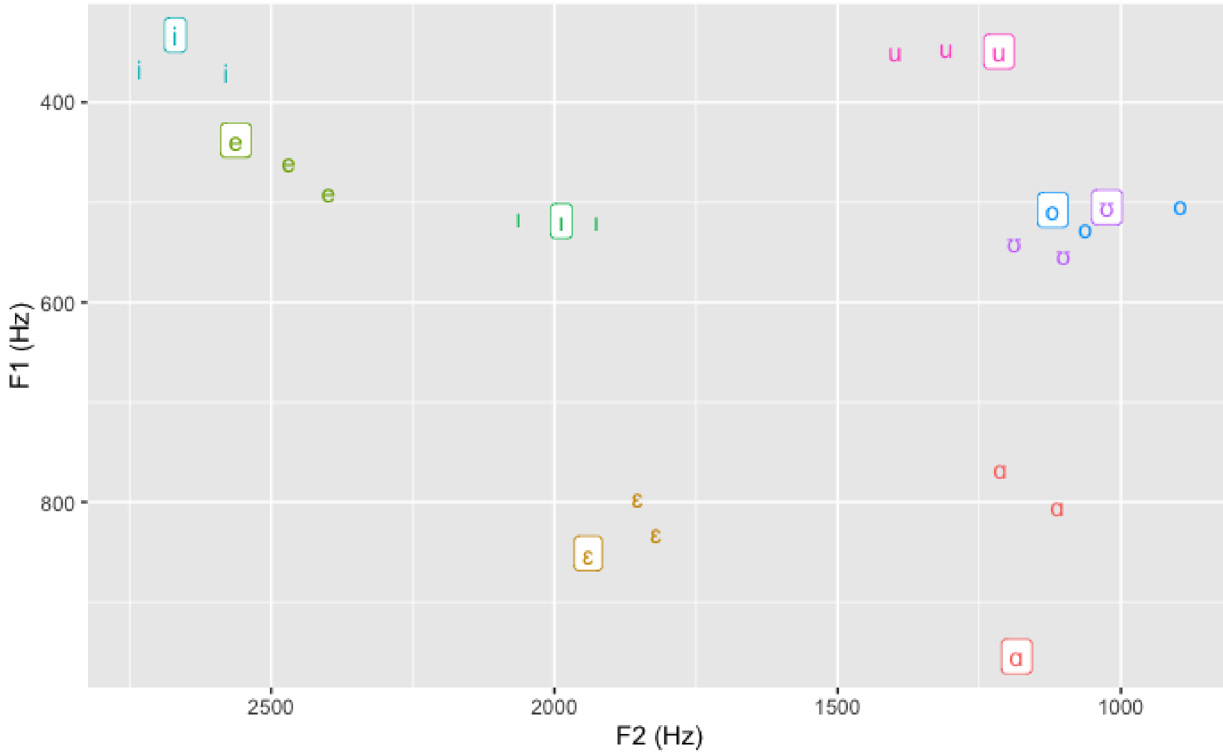

Figure 3.6. Vowel formants of the vowel stimuli for the English AX discrimination experiment, measured at vowel midpoint. The vowel tokens depicted in boxes are the vowel tokens that occurred in the '(phonemically) different vowel' condition. Refer to the text for the duration difference between lax and tense vowels.

The stimulus tokens were concatenated in pairs with a 100 ms silent inter-stimulus interval and a 500 ms silent buffer before the first stimulus token. There were three sets of AX vowel discrimination stimulus pairs: same token stimulus pairs, same vowel different token stimulus pairs, and different vowel stimulus pairs. There were 24 same token stimulus pairs where A and X are acoustically identical stimuli. All 3 tokens of the 8 vowels were combined with themselves in these stimulus pairs. There were 48 same vowel different token stimulus pairs where A and X are phonemically identical but acoustically distinct stimuli. For each of the 8 vowels, the 3 tokens of the vowel were presented in the possible permutations (P(3, 2)), yielding 6 stimulus pairs per vowel. Lastly, there were 56 different vowel stimulus pairs where A and X are phonemically (and hence acoustically) distinct stimuli. For each of the 8 vowels, one token with the least ambiguous vowel formants per vowel was selected and the tokens were presented in the possible permutations (P(8, 2)), yielding 7 stimulus pairs per vowel where the vowel is stimulus A and 7 stimulus pairs per vowel where the vowel is stimulus X.

3.4.1.2.2. ABX vowel discrimination

Stimuli for the ABX vowel discrimination task consisted of 16 monosyllabic /bVb/ nonword tokens with 1 token per each of the English vowels (/ɑ e ɛ i ɪ o u ʊ/) produced by two female native English speakers. The vowel tokens from Speaker 1 were used only in the 'A' and 'B' stimuli in the stimulus triplets and the vowel tokens from Speaker 2 were used only in the 'X' stimuli in the stimulus triplets (Figure 3.7). The vowel tokens of interest were extracted from the audio recordings and the vowel duration, intensity, and pitch contours were manipulated using Praat to control these variables across the stimuli. All lax vowel stimuli (/ɛ ɪ ʊ/) were manipulated to be 100 ms long (mean duration of lax vowels produced by the speakers) and all tense vowel stimuli (/ɑ e i o u/) were manipulated to be 130 ms long (mean duration of the tense vowels produced by the speakers). All vowel stimuli's mean intensity was set to 70 dB and pitch tiers were replaced with a naturally falling pitch tier that was extracted from one of the stimulus tokens. Vowel formants were not manipulated to ensure naturalness of the stimuli.

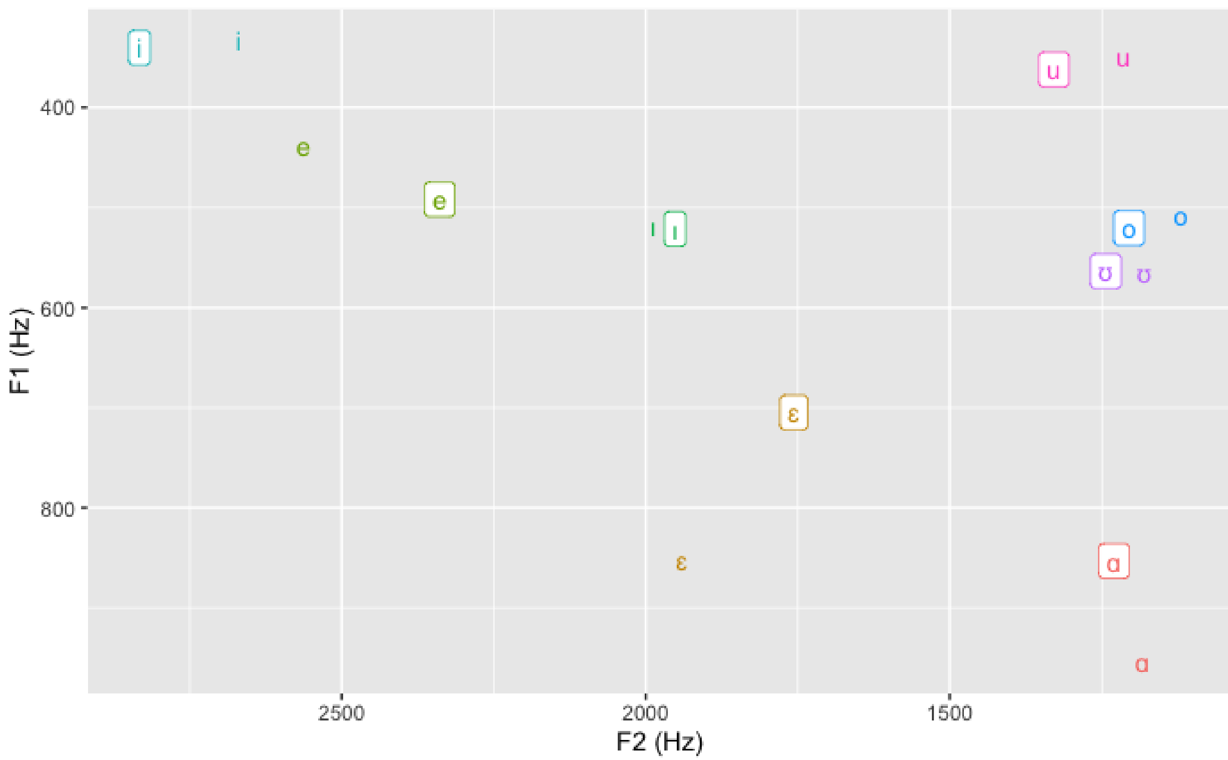

Figure 3.7. Vowel formants of the vowel stimuli for the English ABX vowel discrimination experiment, measured at vowel midpoint. The vowel tokens depicted in boxes are the vowel tokens that were produced by Speaker 1 and that occurred in the A and B positions of the stimulus triplet, and the vowel tokens depicted without boxes are the vowel tokens that were produced by Speaker 2 and that occurred in the X position of the stimulus triplet. Refer to the text for the duration difference between lax and tense vowels.

The stimulus tokens were concatenated in triplets with 500 ms silent inter-stimulus intervals between A and B and between B and X, and a 500 ms silent buffer before the presentation of A in each triplet. There were two sets of ABX vowel discrimination stimulus pairs: A = X triplets and B = X triplets. There were 56 A = X triplets in which the vowel in the target syllable 'X' was identical with the vowel in the first syllable 'A', and 56 B = X triplets in which the vowel in the target syllable 'X' was identical with the vowel in the second syllable 'B'.

3.4.1.3. Procedure

For the participants recruited from Prolific, the vowel perception study took place remotely. The Prolific subject pool was prescreened for residence in the US, English as the first and earliest language in life, English as the primary language, monolingual development, English speaking monolingual status, normal or corrected to normal vision, no hearing difficulties, no cochlear implants, no language related disorders, dyslexia, speech disorders, and ages between 18-65. Participants were informed that the study requires a computer with audio and the use of headphones throughout the experiment, and that the study is not compatible with tablets or mobile devices. Participants were instructed to be in a setting with minimal noise and distractions to begin the experimental procedure and not engage in anything other than the experimental tasks.

For the participants recruited from the Northwestern University Linguistics Subject Pool, the study was conducted in person in Northwestern University's Linguistics Laboratories on the Evanston campus in Illinois. Participants were seated by themselves in soundproof booths in front of a MacBook Air laptop computer and were instructed to wear corded headphones. Participants first gave their informed consent and completed a language background survey, after which they completed the AX vowel discrimination task followed by the ABX vowel discrimination task. The AX and ABX vowel discrimination tasks had 208 and 112 experimental trials, respectively, and procedures were identical to the procedure described above for the Turkish vowel discrimination tasks. In-person participants completed the two vowel perception experiments in a single session and could take breaks during the session with optional break screens halfway through each experiment.

3.4.2. Data analysis and results

Response times (log-transformed) measured from the offset of stimulus X for the accurately responded trials were modeled in a linear mixed effects regression. Trials in which response times were more than 2.5 standard deviations above the mean response time of the respective task were also excluded. The same vowel (dis)similarity measures that were used as predictors in the Turkish vowel discrimination models were calculated for the English stimuli and lexicon. Vowel feature dissimilarity ratio between vowel categories were calculated as the number of features by which the two vowels differ, divided by 4, which is the number of total English vowel features defining the vowels in the stimuli (see Table 3.5). PMI values were calculated over the 61,905 disyllabic English words in the CMU Pronouncing Dictionary. Categorical predictors were sum coded and continuous predictors were centered and standardized.

Table 3.5. Vowel features of the vowels in the English vowel discrimination study.

Vowel	Backness	Rounding	Height	Tenseness
/ɑ/	back	unrounded	low	tense
/ɛ/	front	unrounded	mid	lax
/e/	front	unrounded	mid	tense
/ɪ/	front	unrounded	high	lax
/i/	front	unrounded	high	tense
/o/	back	rounded	mid	tense
/ʊ/	back	rounded	high	lax
/u/	back	rounded	high	tense

3.4.2.1. AX discrimination task

In the AX discrimination task, there was a total of 4592 critical trials and participants responded accurately to 96.82% of the trials. Trials in which an inaccurate response was made and trials in which response times were more than 2.5 standard deviations above the mean response time were excluded (M_{RT} = 970.03 ms, SD_{RT} = 876.52 ms). Response times (log-transformed) from the second stimulus offset for a total of 4446 accurately responded trials were modeled in a linear mixed effects regression (17), paralleling the model structure for the Turkish AX vowel identification model. There were random intercepts by participant and item. No random slopes were included in the model due to singularity and convergence issues.

(17) response time ~ response key
+ Euclidean distance
+ Pillai score
+vowel feature dissimilarity ratio
+ PMI
+ (1 | participant)
+ (1 | item)

None of the fixed effects were significant (Table 3.6). There was a marginally significant main effect of Pillai score (β = -0.02, *SE* = 0.01, t = -1.85, p = 0.07), indicating that participant responses became faster with increasing Pillai score or lesser overlap between the stimulus pair vowel categories although this effect did not reach significance.

Table 3.6. Model results for English AX vowel discrimination task.

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	6.72	0.04	41.98	157.54	< 0.0001***
Response key	0.06	0.04	40.73	1.52	0.14
Euclidean distance	-0.01	0.01	53.32	-0.89	0.38
Pillai score	-0.02	0.01	54.57	-1.85	0.07
Feature dissimilarity ratio	-0.001	0.01	53.73	-0.11	0.91
PMI	0.003	0.01	53.92	0.39	0.70

One way in which stimulus vowels were perceptually discriminated in the English AX vowel discrimination task may have been the duration difference across lax and tense vowels.

Although the lax-tense distinction is coded as one of the features in the feature dissimilarity ratio, the vowel stimuli that differ in their tenseness also differ in their durations, which might have influenced vowel discrimination performance beyond the spectral effects of vowel tenseness. To test this post-hoc hypothesis, I ran a second model where stimulus pair duration mismatch is added as a predictor (-1: duration match between A and X; 1: duration mismatch between A and X) and feature dissimilarity ratio is calculated over 3 vowel features excluding tenseness. This model's results revealed a significant main effect of Pillai score ($\beta = -0.02$, $SE = 0.01$, $t = -2.15$, $p = 0.03$), indicating that participant responses became faster with increasing Pillai score or lesser overlap between the stimulus pair vowel categories (Figure 3.8). None of the other predictors were significant (Table 3.7).

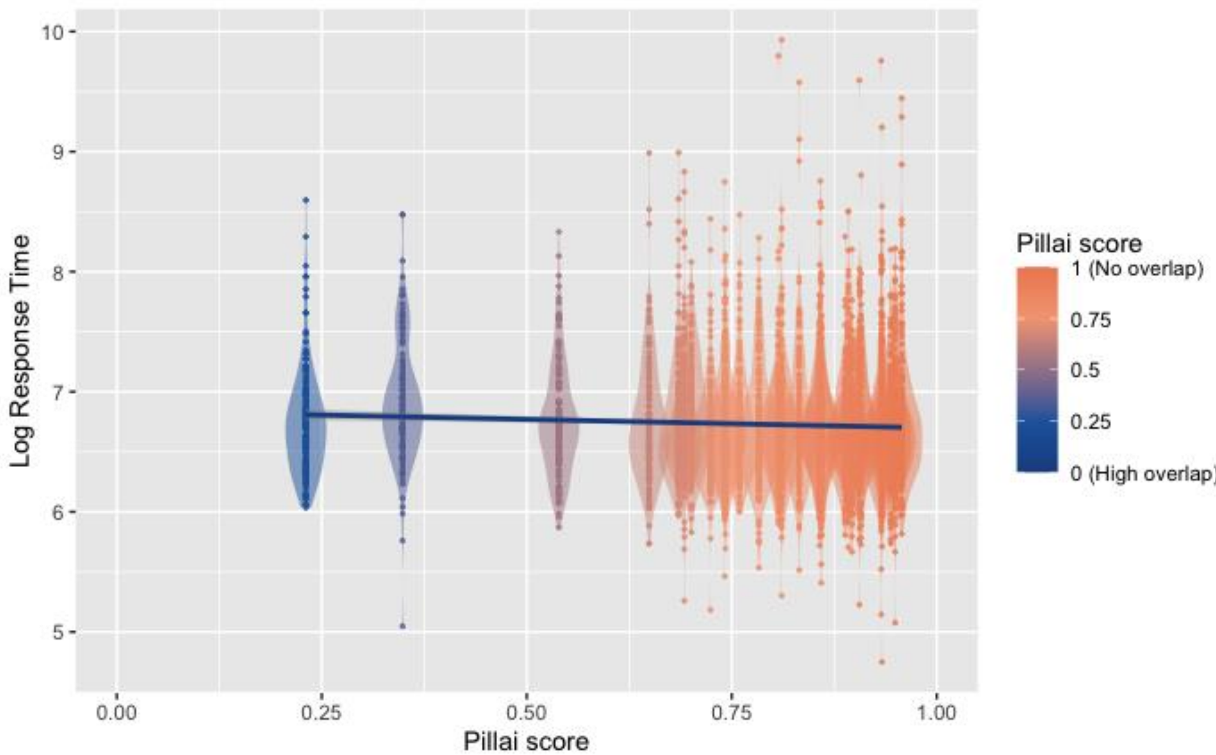

Figure 3.8. Distribution of response times (log transformed) as a function of Pillai scores of stimulus pairs in the English AX vowel discrimination task, with each point representing one trial. Fitted line represents the linear relationship between response time and Pillai score.

Table 3.7. Model 2 results for English AX vowel discrimination task.

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	6.72	0.04	40.79	158.90	< 0.0001***
Response key	0.06	0.04	40.73	1.52	0.14
Euclidean distance	-0.01	0.01	4395.74	-0.68	0.50
Pillai score	-0.02	0.01	4395.87	-2.15	0.03*
Vowel duration mismatch	0.01	0.01	4395.86	11324	0.19
Feature dissimilarity ratio	-0.01	0.01	4395.76	-0.96	0.34
PMI	0.01	0.01	4395.77	0.83	0.41

3.4.2.2. ABX discrimination task

In the ABX discrimination task, there were a total of 4592 trials and participants responded accurately to 88.15% of the trials. Trials in which an inaccurate response was made and trials in which response times were more than 2.5 standard deviations above the mean response time were excluded (M_{RT} = 1308.85 ms, SD_{RT} = 964.90 ms). Response times (log-transformed) from the third stimulus offset for a total of 4048 accurately responded trials were modeled in a linear mixed effects regression (18), paralleling the model structure for the Turkish ABX vowel identification model. There were random intercepts by participant and item as well as random slopes by stimulus order and Euclidean distance by participant. No other random slopes were included in the model due to singularity and convergence issues.

(18) response time ~ stimulus order
+ Euclidean distance
+ Pillai score
+vowel feature dissimilarity ratio
+ PMI
+ (1 + stimulus order + Euclidean distance | participant)
+ (1 | item)

There was a significant main effect of Euclidean distance (β = -0.04, SE = 0.01, t = -3.218, p = 0.002), indicating that as Euclidean distance between two vowel tokens in a stimulus pair increased, response times became shorter (Figure 3.9). None of the remaining fixed effects were significant (Table 3.8).

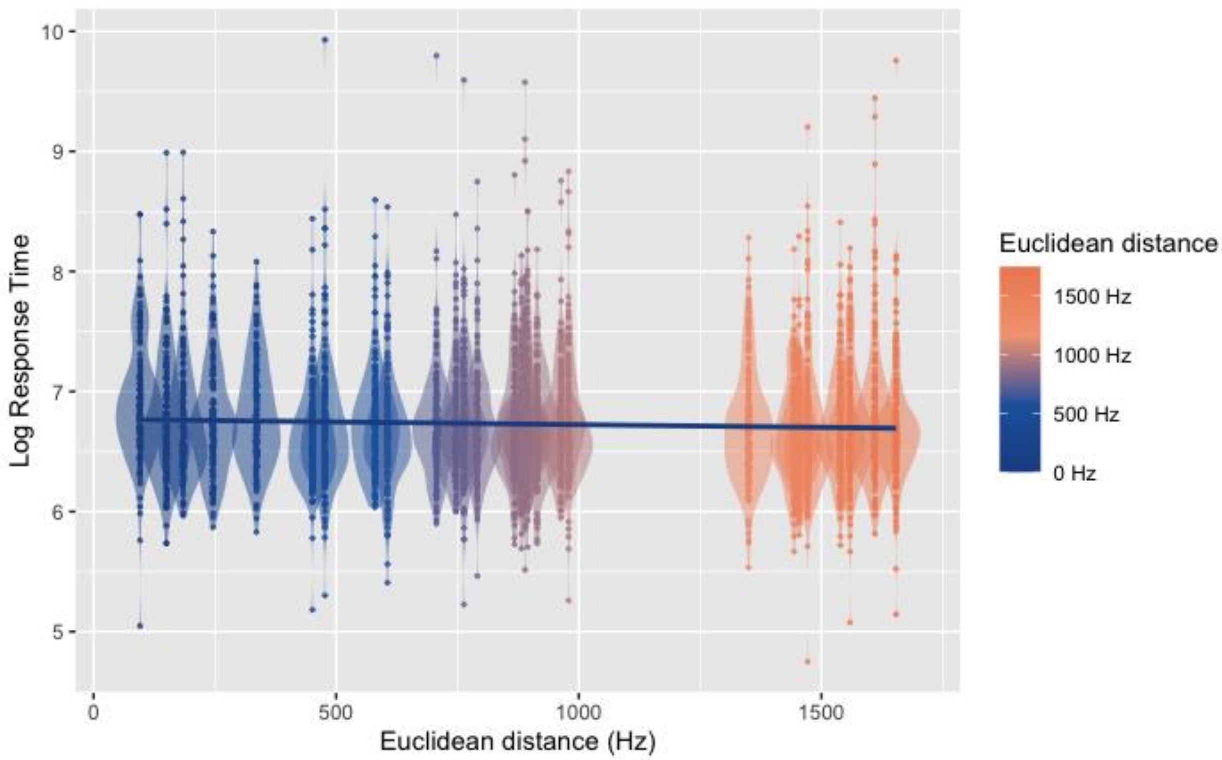

Figure 3.9. Distribution of response times (log transformed) as a function of Euclidean distance between stimulus pairs in the AX vowel discrimination task, with each point representing one trial. Fitted line represents the linear relationship between response time and Euclidean distance.

Table 3.8. Model results for English ABX vowel discrimination task.

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	6.98	0.06	40.79	112.89	< 0.0001***
Stimulus order	-0.003	0.01	35.80	-0.35	0.73
Euclidean distance	-0.04	0.01	57.67	-3.18	0.002**
Pillai score	-0.01	0.01	116.16	-1.08	0.28
Feature dissimilarity ratio	-0.01	0.01	107.40	-0.78	0.44
PMI	-0.01	0.01	106.03	-1.11	0.27

Together, these results suggest that native English listeners' vowel processing invoked vowel categories as evidenced by the near-significant and significant effects of Pillai scores in the AX vowel discrimination models, whereas the ABX vowel discrimination model involved phonetic but not phonological processing as Euclidean distance was the only significant predictor of response times. However, the unexpected pattern of results is also paralleled by the lower accuracy of English listeners in the ABX vowel discrimination task compared to both the AX vowel discrimination task and the accuracy level of Turkish listeners in the ABX vowel discrimination task, suggesting that English listeners found the ABX vowel discrimination task harder. I ran a post-hoc analysis modeling response accuracy in the ABX vowel discrimination task in a binomial logistic regression with stimulus order, Euclidean distance, Pillai score, vowel feature dissimilarity ratio, and PMI as predictors. The results for this model show that higher Pillai scores significantly increased accuracy ($\beta = 0.14$, $SE = 0.06$, $z = 2.57$, $p = 0.01$), indicating that lesser overlap between A and B vowel categories made the stimuli perceptually more dissimilar and easier to discriminate. Stimulus order had a marginally significant effect whereby B = X trials received more accurate responses than A = X trials ($\beta = 0.09$, $SE = 0.05$, $z = 1.92$, $< = 0.055$). Hence, these results suggest that English listeners engaged vowel categories in their processing of vowels, with Pillai scores facilitating accuracy. Within accurately responded trials, the phonetic predictor Euclidean distance facilitated response times. I interpret these results to suggest that English listeners rely on acoustic-phonetic distance and vowel category overlap measures but not vowel features when discriminating between their native vowel categories unlike Turkish listeners.

3.5. General discussion

Turkish and English listeners completed two vowel discrimination tasks in their native languages, an AX vowel discrimination task that is hypothesized to encourage a more heavily acoustic-phonetic vowel processing and an ABX vowel discrimination task that is hypothesized to encourage a more heavily phonological vowel processing due to differences in vowel stimuli, inter-stimulus intervals, and feedback. Response times were the main measure of interest as an indicator of vowel discrimination ease and were modeled with phonetic and phonological measures of vowel (dis)similarity as predictors to investigate the level of processing that listeners were engaged in when performing vowel discrimination. Task demands were manipulated following previous research, and I predicted that both Turkish and English listeners would exhibit faster responses as a function of greater phonetic dissimilarity in the AX vowel discrimination task and as a function of greater phonological dissimilarity in the ABX vowel discrimination task. Moreover, I hypothesized that Turkish listeners would utilize phonological dissimilarity measures involving vowel features and lexical vowel cooccurrence statistics in the ABX vowel discrimination task. The results here partially support these predictions.

To summarize the results (Table 3.9), for Turkish listeners, greater Pillai scores indicating lesser overlap of vowel categories in the F1-F2 space and greater vowel feature dissimilarity ratio predicted faster response times in the AX vowel discrimination task. On the other hand, for English listeners, greater Pillai scores indicating lesser overlap of vowel categories in the F1-F2 space predicted faster responses in the AX vowel discrimination task whereas greater Euclidean distance between stimulus vowel tokens predicted faster responses in the ABX vowel discrimination task. In addition, greater Pillai scores indicating lesser overlap of vowel categories in the F1-F2 space predicted higher accuracy in the English ABX discrimination task. Hence, English listeners engaged in similar levels of phonetic and phonological processing in both AX and ABX vowel discrimination tasks, with more evidence of a predictor invoking vowel categories influencing vowel perception in the AX vowel discrimination task. Contrary to predictions from previous research, phonological predictors were not more influential in the ABX vowel discrimination tasks. As such, the results suggest that the task demand manipulations to target phonetic or phonological processing selectively in the AX and ABX vowel discrimination tasks respectively were not successful.

Table 3.9. Summary of the significant effects in the vowel discrimination experiments.

Experiment	Significant predictor	Predictor type: Phonetic	Predictor type: Phonological
Turkish AX	Pillai (less overlap → faster RT)	✓	✓
English AX	Pillai (less overlap → faster RT)	✓	✓
English ABX	Euclidean distance (greater distance → faster RT)	✓	
English ABX	Pillai (less overlap → higher accuracy)	✓	✓

The motivation behind the cross-linguistic comparison was to investigate whether listeners experienced with harmony and non-harmony languages alike invoked vowel features in vowel discrimination. Although English listeners' response times in the ABX vowel discrimination task did not reveal any significant phonological predictors, that vowel feature dissimilarity ratio was a significant predictor in AX vowel discrimination for Turkish but not English listeners is in support of the hypothesis that vowel features are a prominent part of perceived vowel similarity in Turkish but not necessarily in other languages. However, as it is not clear whether English listeners engaged in phonological processing in the ABX vowel discrimination task in a manner that is comparable to Turkish listeners, cross-linguistic comparisons with other harmony and non-harmony languages or with other methods of probing perceived vowel similarity is necessary to reach a definitive answer.

Lastly, my research question about the hypothesized relationship between lexical vowel cooccurrence statistics and vowel features in Turkish remains unanswered. Despite the results here suggesting that the AX vowel discrimination task involved phonetic and phonological processing, I do not find evidence that vowel cooccurrence relationships as they manifest in the lexicon are engaged in this task. On the other hand, Turkish listeners were found to invoke vowel features in the AX vowel discrimination task. My hypothesis about the relationship between vowel cooccurrence in the lexicon and vowel features in harmony languages is one in which the latter emerges from the former, and hence where vowel features are invoked, I hypothesize they are invoked via vowel coactivation based on vowel cooccurrence. As such, that vowel features were part of the Turkish vowel perception mechanism without engaging vowel cooccurrence contradicts my hypothesis. In the next chapter, I present vowel confusion data from a vowel identification study in Turkish, which is predicted to engage vowel categories and vowel phonology more directly than vowel discrimination tasks reported in this chapter, and hence might be more suitable for exploring effects of vowel cooccurrence on perceived vowel similarity.

CHAPTER 4

VOWEL IDENTIFICATION EXPERIMENT

This study aims to examine whether, as predicted by exemplar models (e.g., Pierrehumbert, 2001), greater likelihood of vowel cooccurrence leads to greater perceived vowel similarity beyond the phonetic similarity of vowels in Turkish. In the previous chapter, I presented two vowel discrimination experiments that were designed to target varying levels of acoustic-phonetic and phonological processing of vowels, thereby aiming to reveal the relative contributions of acoustic-phonetic and phonological determiners of perceived vowel similarity, including likelihood of vowel cooccurrence. As an extension of the vowel discrimination experiments and to more directly target phonological processing involved in perceived vowel similarity, I report results from a vowel identification experiment in the present chapter.

To anticipate the results, I find a significant effect of likelihood of vowel cooccurrence on perceived vowel similarity, although this effect is not in the expected direction based on my hypothesis that higher likelihood of vowel cooccurrence leads to higher perceived vowel similarity. However, an unexpected finding that whether a stimulus vowel has relatively extreme F1-F2 values and hence is more peripherally situated in the vowel space is a significant predictor of perceptual confusion rates invites considering vowel peripherality as a predictor of perceived vowel similarity beyond the factors that were included in the model originally. Following this observation, I present a post-hoc analysis with vowel peripherality as a predictor of perceptual confusions. I discuss how these findings relate to previous claims about Turkish vowel harmony being perceptually beneficial for nonperipheral vowels with less distinct F1-F2 values (e.g., (Clements & Sezer, 1982; Kaun, 1995; Suomi, 1983).

4.1. Vowel identification as a measure of perceived vowel similarity

The set of experiments in this work are designed to complement each other towards exploring the phonetic and phonological predictors of perceived vowel similarity, and in particular, how the likelihood of vowel cooccurrence in the lexicon influences perceived vowel similarity. From an exemplar models perspective, vowels that are coactivated by their phonetic similarity form a phoneme category (e.g., Pierrehumbert, 2001). Similarly, on the level of lexical representations, vowels that are coactivated by cooccurring in a lexical item can form a harmony class (Baker, 2009; Cole, 2009). Hence, a prediction of exemplar models is that vowels that are highly likely to cooccur in the lexicon will be perceived to be more similar than expected based on their phonetic similarity alone. To test this prediction, an experimental paradigm that targets a level of processing beyond the acoustic-phonetic level is necessary.

The studies reviewed in Chapter 3 on acoustic-phonetic as opposed to phonological processing of vowels argued that tasks that require phonemically labeling vowels engage phonological processing more than vowel discrimination tasks that do not rely on phoneme labels: the finding of 'continuous vowel perception' was based on listeners being able to discriminate steps from a vowel continuum more granularly than their identification functions of the same stimuli predicted. In a similar vein, it can be argued that phonemic vowel identification is inherently a more phonological task than is vowel discrimination. In this chapter, I report a Turkish vowel identification study where native Turkish listeners were asked to phonemically identify vowel stimuli they heard in an 8-alternative forced choice paradigm where each stimulus vowel token had to be categorized as one of the 8 canonical Turkish vowels. Inaccurate vowel identification responses are taken to result from perceptual confusion indicating that a stimulus vowel and the response vowel are perceived to be similar.

I model the vowel identification responses with phonetic, phonological, and phonotactic measures of stimulus-response vowel similarity to examine whether a higher likelihood of vowel cooccurrence in the lexicon leads to higher perceptual confusability beyond what would be expected from the phonetic similarity of stimulus-response vowel pairs. I compare Turkish vowel identification with English vowel identification data from Hillenbrand et al., (1995) to compare the effect of likelihood of vowel cooccurrence on perceived vowel similarity across harmony and nonharmony languages. To anticipate the results, although I do not find a general pattern whereby vowels that are more likely to cooccur in the same word are more likely to be perceptually confused, I find an interaction between whether a stimulus vowel is more peripherally situated in the vowel space and the effect of likelihood of vowel cooccurrence on perceived vowel similarity. In the upcoming sections, I describe my approach to modeling vowel identification responses as a measure of perceived vowel similarity with phonetic and phonological predictors, followed by a recap of previous accounts of Turkish vowel harmony that justify considering vowel peripherality as an additional predictor.

4.1.1. Modeling vowel identification (Hall and Hume, 2015)

A vowel identification study by Hall and Hume (2015) models perceptual confusions in French vowel identification to examine how acoustic-phonetic properties of the stimuli and the lexical contrastiveness of vowels influence perceived vowel similarity. In Hall and Hume's vowel identification task, native French listeners were auditorily presented with French vowel stimuli. Inaccurate vowel identification responses were treated as perceptual confusions and therefore as a measure of perceived vowel similarity. The vowels that were most frequently confused with other vowels were schwa, /œ/ and /ø/, which were accurately identified in less than 40% of the trials.

These vowels were almost entirely confused with each other and had similar F1-F2 values. In contrast, the vowels that were least frequently confused with other vowels were /i/, /y/ and /a/, which were accurately identified as the intended vowel in more than 99% of the trials. These vowels are peripherally situated in the French vowel space: they have relatively extreme F1 and/or F2 values that set them apart from other vowels, although /u/, despite being a peripheral vowel with distinct F1-F2 values, was accurately identified in only 78% of the trials. Confusions across rounded and unrounded vowels were rare, such as /i/ and /y/, despite their overlapping F1-F2 values, and confusions with a vowel's open or close counterpart were common within mid vowels. Thus, the patterns of vowel confusion suggested that there were potentially other factors influencing perceived vowel similarity beyond phonetic similarity in the F1-F2 space.

To identify the factors contributing to perceived vowel similarity and their respective roles, Hall and Hume (2015) modeled the full set of vowel confusion data (quantified as percentages of given responses for every stimulus vowel) with phonetic and phonological measures of vowel similarity as predictors using linear regression. Phonetic predictors included in the model were the Euclidean distance and the vowel duration difference between stimulus and response vowels. Phonological predictors were gradient, usage-based measures of contrastiveness: *Functional load*, which conceptualizes how many words of the lexicon are distinguished by a pair of sounds, with high functional load distinguishing many words, indicating greater contrastiveness, and *degree of complementary distribution*, which captures the degree to which the distributions of two sounds across contexts in the lexicon can be characterized as contrastive as opposed to allophonic.

The French vowel identification stimuli in Hall and Hume (2015) contained 14 French vowels, all of which are phonemically contrastive in French yet vary in how contrastive their distributions are in the lexicon as measured by functional load and degree of complementary distribution. Functional load was quantified in two measures: the number of minimal word pairs that distinguish each vowel pair, and the change in total entropy or uncertainty of the full sound system when two vowels are merged, neutralizing the phonological contrast. The degree of complementary distribution was quantified as the predictability of vowel A as opposed to vowel B across all the contexts that A and B occur in. Lower predictability indicates that the two vowels have overlapping distributions and are more contrastive, whereas higher predictability indicates that the two vowels occur in distinct contexts and are more allophonic. The degrees of complementary distribution for each of the vowel pairs ranged from almost perfect predictability (such as for /œ/ and /ø/) to almost perfect unpredictability (such as for /a/ and /i/). Lastly, vowel frequency ratios of vowel pairs were also added as a predictor to the model to account for a hypothesized vowel identification bias towards the relatively more frequent vowel of each vowel pair. The phonological predictors and the vowel frequency ratios were calculated from the French corpus Lexique, and both type frequency-based and token frequency-based calculations were carried out to conduct comparative analyses.

Hall and Hume (2015) modeled percent confusability of each vowel pair in linear regressions with Euclidean distance, duration difference, measures of functional load, degree of complementary distribution, frequency ratio, and their 2- and 3-way interactions as predictors. All of the factors had effects in the expected direction, although duration difference and frequency ratio were not significant and were dropped from the final models. Model comparisons for the remaining factors revealed that Euclidean distance was the strongest predictor, with increasing phonetic distance between stimulus and response vowels leading to lower vowel confusability. The second strongest predictor was degree of complementary distribution such that lower degree of complementary distribution, indicating that the stimulus and response vowels are more contrastive with one another, predicted lower vowel confusability. Lastly, greater functional load was also associated with lower vowel confusability.

Together with the significant interactions, the final model explained 79% of the variance in the vowel confusability data. The final model displayed a large improvement from a model that only included Euclidean distance and duration difference, which accounted for 36% of the variance, suggesting that the phonological predictors of similarity examined in the study, such as functional load and degree of complementary distribution, played a role in the perceived similarity of vowels. In addition, type frequency-based phonological measures of similarity were stronger predictors of perceived vowel similarity than token frequency-based measures. The authors hence argue that the phonological component of perceived similarity is emergent from word-level units of the lexicon rather than from segmental units or perceived exemplars.

Hall and Hume's (2015) analysis shows that perceptual confusions in vowel identification can be modeled with phonetic and phonological measures of vowel similarity as predictors to examine the relative contribution of each predictor to perceived vowel similarity. I follow their direction in my own analysis of Turkish vowel identification data where I ask questions about likelihood of vowel cooccurrence as a predictor of perceived vowel similarity rather than the contrastiveness of vowels. I utilize Poisson regressions with phonetic and phonological-phonotactic predictors to model perceived model similarity. Based on my hypothesis that higher likelihood of vowel cooccurrence in the lexicon leads to higher perceived vowel similarity, I predict higher rates of perceptual confusion in Turkish vowel identification between vowels that are more likely to cooccur. For comparison with Turkish, I also model English vowel identification data from Hillenbrand et al. (1995). I predict that the lack of vowel harmony in English will make the effect of likelihood of vowel cooccurrence weaker compared to Turkish.

4.2. Turkish vowel identification

4.2.1. Methods

4.2.1.1. Participants

Participants were 40 monolingually raised native speakers of Turkish aged 18-61 (M_{age} = 26.35; 24 women, 1 other) residing in Istanbul, Turkey at the time of the study. Thirty-five participants self-reported English as a second language, and 32 of these participants self-reported they were fluent in English. One participant self-reported German as a second language they were fluent in. Participants' self-reported age of onset for English was ranged from 5 to 21 (M_{age} = 9.47). Sixteen participants self-reported knowing a third language, 7 participants self-reported knowing a fourth language, and 1 participant self-reported knowing a fifth language. More specifically, 6 participants knew Arabic, 5 participants knew French, 4 participants knew German, 3 participants knew Korean, 3 participants knew Japanese, 1 participant knew Spanish, 1 participant knew Chinese, and 1 participant knew Indonesian. Participants were paid ₺300 Turkish Liras for their participation at the completion of the experiment.

4.2.1.2. Stimuli

Stimuli consisted of one monosyllabic /bVb/ nonword token per each of the 8 Turkish vowels (/ɑ e ɯ i o œ u y/) produced by 20 speakers from the Turkish vowel production study (see Chapter 2), totaling 160 stimulus tokens. Stimulus selection involved examining the vowel formants for all monosyllabic nonword productions across 20 speakers to identify a set of tokens that represented the variability and spread of each of the Turkish monosyllabic nonword vowel clouds (Figure 4.1). The naturally produced monosyllabic nonword tokens were excised from the audio recordings using a Praat script that extracted the selected vowel tokens according to the audited vowel boundaries set by MFA by adding the 60 ms preceding the vowel onset and the 100 ms following the vowel offset to include the syllable onset and coda. All stimulus tokens' mean amplitude was equalized at 70 dB and a 440 ms silent buffer was added before the syllable onset. Vowel formants, duration, and pitch were not manipulated to preserve the naturalness of the stimuli.

Figure 4.1. F1-F2 formant values of the stimulus vowel tokens in the Turkish vowel identification task.

4.2.1.3. Procedure

The vowel perception study was conducted in person in reserved rooms on school campuses in Istanbul, Turkey. Participants were seated in front of a MacBook Air laptop computer with a touchpad and wore corded headphones. Participants first gave their informed consent and completed a language background survey, after which they completed the vowel perception study consisting of the vowel identification task, an AX vowel discrimination task and lastly an ABX vowel discrimination task (see Chapter 3 for details). Participants completed the three vowel perception experiments in a single session and could take breaks during the session with optional break screens halfway through each experiment.

The vowel identification task was an 8-alternative forced choice (8AFC) task consisting of 160 trials. Each stimulus token was presented once. Each trial began with a play button visually presented on screen. Upon clicking the play button by bringing the cursor on it using the touchpad, a randomly selected stimulus token was auditorily presented along with a loudspeaker icon visually presented on screen. Following the auditory presentation of the stimulus token, the 8 response options were visually presented on screen in a circular array of 4 rows of tiles (see Figure 4.2) and participants were asked to indicate which vowel they heard by clicking on the corresponding response option. The Turkish orthography is transparent in its representation of the 8 phonemic vowels presented in the monosyllabic nonword stimuli, and the orthographic representations of the Turkish vowels in small caps were used to label the response options (/ɑ/: ‘a’, /e/: ‘e’, /ɯ/: ‘ı’, /i/: ‘i’, /o/: ‘o’, /œ/: ‘ö’, /u/: ‘u’, /y/: ‘ü’). To facilitate the search for the response vowel, the response options were presented in alphabetical order starting from the top left tile in each trial. As such, the four back vowels were presented on the left half of the circular array of tiles and the four front vowels were presented on the right half of the circular array of tiles, and in each row, the response options had vowel pairs that only contrasted in backness. In each trial, the stimulus token was played only once. There was no time constraint for making a response and the trial ended upon participant response, after which a new trial began with the play button which forced participants to move their cursor back to the center of the circular array at the beginning of each trial. Participants were instructed to pick the response that seems best if they are not sure. The vowel identification task began with 4 practice trials. A screen announcing an optional break was presented halfway through the task.

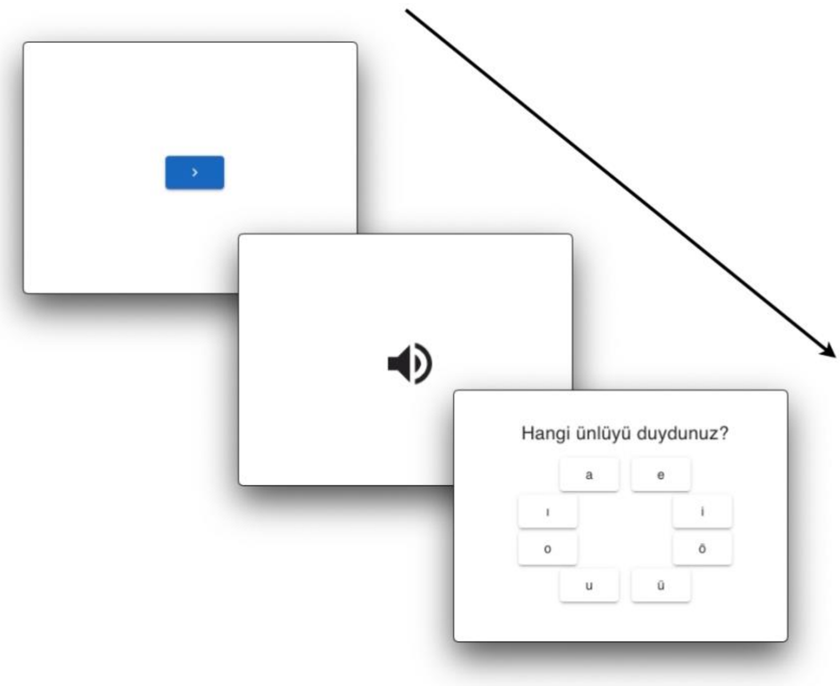

Figure 4.2. Flow of a trial in the 8AFC Turkish vowel identification task. Refer to the text for details.

4.2.2. Results

A total of 6400 vowel identification responses were collected and analyzed, with the inaccurate responses interpreted as perceptual confusions between a stimulus vowel and the response vowel, and with the rate of perceptual confusion per vowel pair interpreted as a measure of perceived vowel similarity in Turkish. The overall accuracy rate was 89.53%, with a total of 670 inaccurate vowel identifications (see Table 4.1), i.e., perceptual confusions. Figure 4.3 depicts the empirical data as a heatmap of participant response proportions to each stimulus vowel category. In the next section, I provide an exploratory overview of response times and accuracy in vowel identification before presenting the hypothesis-driven analyses of the vowel identification responses.

Table 4.1. Response counts in the Turkish vowel identification task.

	Response vowel							
Stimulus vowel	/ɑ/	/ɯ/	/o/	/u/	/e/	/i/	/œ/	/y/
/ɑ/	773		25		2			
/ɯ/		559	1	234				6
/o/	19	4	744	33				
/u/		19	31	750				
/e/		19			712	22	44	3
/i/		1				792	1	6
/œ/	7	50	2	3	26		691	21
/y/		44		1	2	24	20	709

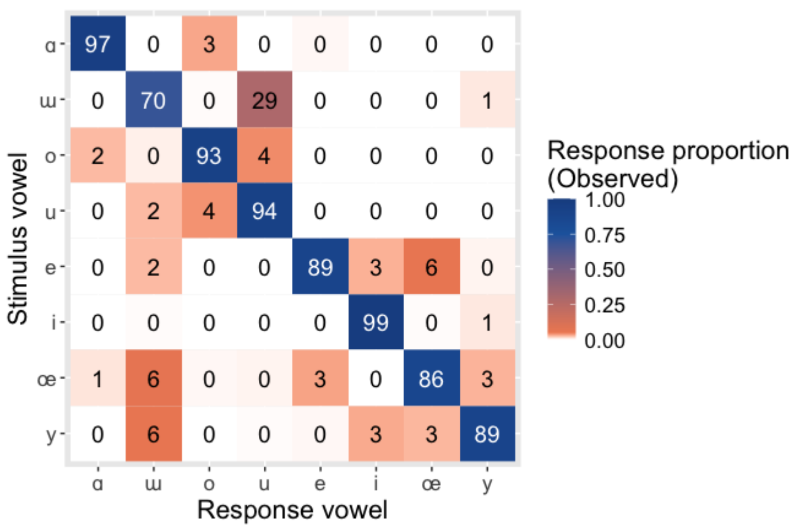

Figure 4.3. Heatmap depicting the observed response proportions in the 8AFC Turkish vowel identification task. Response proportions calculated over each stimulus vowel category are represented as response percentages in each cell. Shaded cells represent responses that were observed in the data with the accurate responses in the diagonal and cell shading representing response proportion per stimulus vowel category. Unshaded (white) cells represent responses that were not observed in the data.

4.2.2.1. Response time and vowel identification accuracy

A linear mixed effects regression modeling response time with accuracy as the predictor and random intercepts by participant and item revealed that accurate vowel identification responses were faster than inaccurate vowel identification responses (β = -196.18, *SE* = 12.68, *df* = 4474.44, *t* = -15.47, $p < 0.0001$), suggesting that in trials that received inaccurate responses, listeners might have perceived the stimuli to be harder to identify and phonemically ambiguous. A binomial logistic regression modeling vowel identification accuracy with stimulus vowel category as the predictor (with /ɑ/ as the reference level) and random intercepts by participant and item revealed that accuracy rate varied by stimulus vowel category, with the vowel categories /ɯ œ y/ exhibiting significantly lower accuracy rates compared to /ɑ/ (/ɯ/: β = -3.30, *SE* = 0.61, *z* = -5.43, $p < 0.0001$; /œ/: β = -2.15, *SE* = 0.61, *z* = -3.51, *p* = 0.0005; /y/: β = -1.77, *SE* = 0.62, *z* = -2.87, *p* = 0.004). The vowel category /e/ was marginally less accurate compared to /ɑ/ (β = -1.23, *SE* = 0.63, *z* = -1.94, *p* = 0.052), and the vowel category /i/ was marginally more accurate compared to /ɑ/ (β = 1.37, *SE* = 0.76, *z* = 1.81, *p* = 0.07).

The vowels that had significantly lower vowel identification accuracy (i.e., /ɯ œ y/; cf. Yılmaz, 2022) have relatively less distinct F1-F2 values and occupy the more crowded and central part of the Turkish vowel space (see Chapter 2) as opposed to being more peripherally situated with relatively more extreme F1-F2 values. The French vowel identification accuracy results in Hall and Hume (2015) had a similar yet less clear pattern where, in general, less peripheral vowels were identified less accurately compared to vowels with more extreme F1-F2 values. The pattern of results here can be attributed to vowel phonetics: A general prediction concerning perceptual confusions in vowel identification is that vowels that are phonetically dissimilar and distant in the vowel space are predicted to be perceptually confused with each other at a lower rate than vowels that are more similar and closer to one another in the vowel space. Accordingly, given its relatively extreme F1-F2 values and its distance from other vowels along both F1 and F2 dimensions, a vowel such as /ɑ/ is predicted to exhibit fewer perceptual confusions and hence be identified with higher accuracy than a vowel such as /œ/.

As illustrated above, the Turkish vowels /ɯ œ y/ that are identified with significantly lower accuracy arguably diverge from vowels such as /ɑ i u/ in terms of their relative phonetic distinctiveness. However, recall that Turkish linguists have also proposed phonological differences between /ɯ œ y/ and /ɑ e i o u/ in Turkish (e.g., Clements & Sezer, 1982; Suomi, 1983; see Chapter 1). Suomi argues that /ɯ œ y/ are phonetically and perceptually weak vowels in Turkish: these vowels have neighboring strong vowels with more extreme F2 values on their F1 level, and hence are perceptually disadvantaged. According to Suomi, Turkish vowel harmony is perceptually motivated to facilitate the accurate identification of /ɯ œ y/ by increasing their contextual predictability via vowel cooccurrence restrictions. Converging with Suomi, Clements and Sezer argue that disharmonic Turkish vowels are more likely to cooccur if the vowels are from the vowel set /ɑ e i o u/, whereas vowels from the vowel set /ɯ œ y/ are subject to a greater degree of cooccurrence restrictions. My findings in Chapter 1 support Clements and Sezer's claim. These observations lead me to consider 'vowel peripherality' (see Chapter 1) as a phonetic factor that might predict not only the vowel identification accuracy rates of individual vowels as reported above but also vowel cooccurrence patterns in Turkish and by extension perceived similarity of Turkish vowel pairs. In the upcoming sections, I first present the results from the preplanned analyses and then discuss how vowel peripherality might provide further insight in interpreting the results. I conclude by presenting a post-hoc analysis where I add vowel peripherality as an additional predictor.

4.2.2.2. Perceived vowel similarity

My goal in the present vowel identification experiment is to examine the predictors of perceived vowel similarity and in particular the likelihood of vowel cooccurrence as a predictor.

To model the perceptual confusion rates, I calculated the number of times each stimulus vowel token was identified as each one of the response vowels. I modeled the resulting count data in a Poisson regression.

As the count data collapses across individual responses per each item-response pair, there can be no by-participant or by-item random effects in the model. The loss of random effects structure in the data is dispreferred, as the vowel identification experiment involved multiple measurements, both by participant and by item. One modeling alternative that would preserve the random effects structure is multinomial logistic regression, so I also modeled the full set of individual vowel identification responses in sequential multinomial logistic regressions that added the predictors one by one. The simplest model with only stimulus vowel token F1 and F2 as predictors did not converge when there were by-participant random intercepts, and as a result all subsequent models only had by-item random intercepts. Yet, the most complex model where Pointwise Mutual Information (PMI) as a measure of phonotactic similarity was added as a predictor did not converge unless the by-item random intercepts were also excluded. As the most complex multinomial logistic regression model converged only without random intercepts or slopes, multinomial logistic regression modeling had no advantage over Poisson regression in terms of random effects structure. Moreover, Poisson regression results have a more straightforward interpretation as the dependent variable is response counts. In contrast, multinomial logistic regression coefficients denote predicted response probability relative to a reference response vowel category. For these reasons and for the sake of brevity, I only report the results from the Poisson regression in this chapter.

To prepare the data for the Poisson regression analysis, the vowel identification data was converted into count data representing how many times each stimulus vowel token was identified as each of the 8 response vowel categories. The stimuli were presented once to each of the 40 participants. Thus, each of the 160 stimulus vowel tokens received 40 vowel identification responses. For each stimulus token, the number of times each response could be made ranged from 0 to 40, with a response count of 0 indicating that that response was not observed in the data. Each individual response can be of two types: an accurate identification and a perceptual confusion (Figure 4.4). The response type of interest here is a perceptual confusion, as accurate vowel identification responses are not informative about perceived vowel similarity. Hence, accurate vowel identification response counts were excluded from the count data and the remaining 1120 perceptual confusion response counts (160 stimuli x 7 inaccurate response vowel categories) were modeled in the Poisson regression. In this reduced dataset, a higher response count indicates higher perceived similarity. Half of the response counts were in the 1-2 range.

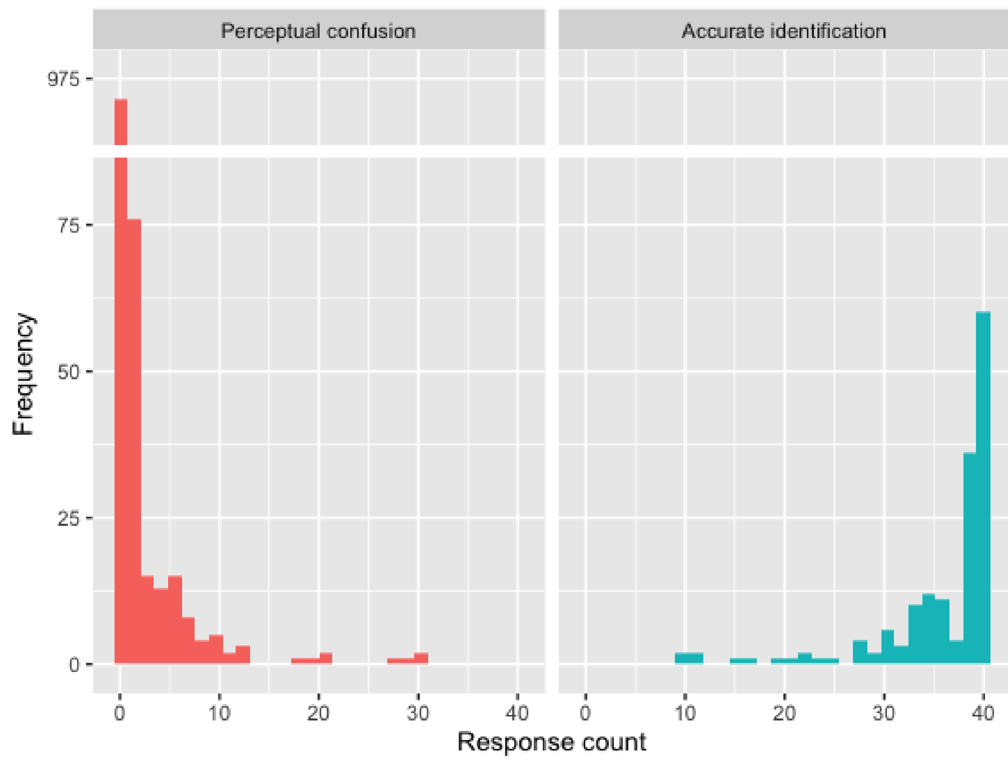

Figure 4.4. Histogram of response counts by stimulus vowel token.

A closer inspection of the count data reveals an excess of response counts that were 0. To be exact, 28 out of 56 possible perceptual confusion responses were not observed, which comprises 75.86% of the perceptual confusion count data. The large number of 0 counts also led to overdispersion in the data with higher variance (σ^2 = 6.64) than mean perceptual confusion count (*M* = 0.60). To handle the excess of 0 counts and the overdispersion in the data, I opted for a zero-inflated Poisson regression (`pscl` package; Jackman, 2024; Zeileis at al., 2008).

Zero-inflated Poisson regression consists of a Poisson regression and a binomial logistic regression. Poisson regression is a type of regression that models variation in count data, yielding predicted counts that are part of a count distribution with 0 as the lower bound and no upper bound. As such, Poisson regressions can predict 0 counts. However, in a zero-inflated Poisson regression, observed 0 counts are modeled under the assumption that there are two types of processes that generate 0 counts in the data: one that generates 0 counts and non-zero counts alike and one that generates 0 counts only. Poisson regression can model 0 counts that are of the former type whereas binomial logistic regression can model the 0 counts of the latter type. Binomial logistic regression models the 0 counts in the data in terms of whether a 0 count is more likely to be observed as part of the count distribution and hence predicted as part of the Poisson regression or not. In other words, the binomial logistic regression indicates whether an observed 0 would be predicted to have a positive value other than 0 or would always be predicted to be a 0 if the predictors had different values. As a result, part of the large number of observed 0 counts are accounted for by the binomial logistic regression, thereby reducing the skew caused by the 0 counts in the count distribution modeled in the Poisson regression.

Following the modeling logic set forth in the vowel discrimination models in Chapter 3, I modeled the vowel identification count data as a function of three sets of phonetic and phonological predictors (Table 4.2). As noted in the preceding chapters, the first set of predictors are the phonetic descriptors of the stimulus vowel token, which represent information processed as part of the early phonetic processing of the speech signal. The second and third set of predictors both refer to vowel categories and hence involve phonological processing. However, the predictors of the second set are arguably more phonetic relative to the third set of predictors, which include PMI as a measure of cooccurrence similarity as calculated over 18,166 disyllabic words in the Turkish lexicon.

Table 4.2. Predictors in the zero-inflated Poisson regression modeling the vowel identification count data.

Predictor type	**Predictor**	**Notes**
Phonetic descriptor of the stimulus vowel token	Stimulus vowel token F1 Stimulus vowel token F2 Speaker gender Stimulus vowel token duration	
	Stimulus vowel token F1 : speaker gender	Interaction term (dropped due to collinearity)
	Stimulus vowel token F2 : speaker gender	Interaction term
Phonetic similarity between the stimulus and the response vowel category	Euclidean distance	Measure of phonetic distance between stimulus token and response vowel category's centroid in F1-F2 space; higher scores indicate greater distance
	Pillai score	Measure of phonetic overlap between stimulus and response vowel categories in F1-F2 space; higher scores indicate lesser overlap
Phonological similarity between the stimulus and the response vowel category	Feature dissimilarity ratio	Measure of phonological feature dissimilarity between stimulus and response vowel categories over the 3 Turkish vowel features [high, back, round]; higher scores indicate fewer shared features
	PMI	Pairwise mutual information – a measure of likelihood of cooccurrence between stimulus and response vowel categories; higher scores indicate higher likelihood of vowel cooccurrence

The same set of predictors were included in both the logistic regression and the Poisson regression parts of the model. I predict that higher phonetic and phonological dissimilarity as measured by higher Euclidean distance, higher Pillai score, higher feature dissimilarity ratio, and lower likelihood of vowel cooccurrence will be reflected in higher perceived dissimilarity and hence yield no perceptual confusions at all or fewer perceptual confusions. A significant positive coefficient in the binomial logistic regression suggests higher likelihood of observing no perceptual confusions at all. A significant negative coefficient in the Poisson regression suggests fewer predicted perceptual confusions. I report the results from the logistic regression in Table 4.3 and the results from the Poisson regression in Table 4.4.

Table 4.3. Model results from the logistic regression modeling the 0 counts.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	2.15	0.29	7.50	< 0.001***
Stimulus vowel token F1	-0.13	0.17	-0.75	0.45
Stimulus vowel token F2	-0.73	0.24	-3.06	0.002**
Speaker gender	0.16	0.24	0.66	0.51
Stimulus vowel token duration	-0.01	0.14	-0.07	0.94
Euclidean distance	1.66	0.30	5.46	< 0.001***
Pillai score	1.03	0.25	4.15	< 0.001***
Feature dissimilarity ratio	0.56	0.15	3.69	< 0.001***
PMI	0.41	0.14	3.04	0.002**
Stimulus vowel token F2 : speaker gender	-0.04	0.27	-0.13	0.89

The results from the logistic regression suggest that there is a significant negative relationship between vowel F2 and perceptual confusions with 0 counts, suggesting that stimulus vowels with higher F2 are more likely to have 0 counts as part of the count distribution. Euclidean distance, Pillai score, and vowel feature dissimilarity ratio between stimulus vowel and response vowel category all had significant and positive main effects indicating that as phonetic or phonological distance between two vowels increases, the likelihood of observing no perceptual confusions at all increases. Lastly, there was a significant and unexpectedly positive main effect of PMI, suggesting that as the likelihood of two vowels cooccurring increases, the likelihood of observing no perceptual confusions at all increases.

Table 4.4. Model results from the Poisson regression modeling the count data.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.01	0.15	0.05	0.96
Stimulus vowel token F1	-0.02	0.09	-0.23	0.82
Stimulus vowel token F2	0.47	0.13	3.68	< 0.001***
Speaker gender	0.09	0.10	0.93	0.35
Stimulus vowel token duration	-0.29	0.06	-5.16	< 0.001***
Euclidean distance	-1.38	0.15	-9.37	< 0.001***
Pillai score	-0.11	0.12	-0.91	0.36
Feature dissimilarity ratio	-0.34	0.07	-4.93	< 0.001***
PMI	-0.22	0.06	-3.68	< 0.001***
Stimulus vowel token F2 : speaker gender	-0.07	0.14	-0.52	0.60

The results from the Poisson regression were mainly in line with the results from the logistic regression with most significant effects in opposite directions. Note that a Poisson regression yields coefficients that are logarithms rather than the exact predicted counts. Stimulus vowel token F2 had a significant and positive main effect indicating that stimuli with higher F2 are more likely to be perceptually confused. Stimulus vowel token duration had a significant and negative main effect indicating that stimuli with longer duration were less likely to be perceptually confused than stimuli with shorter duration. There were significant and negative main effects of Euclidean distance and vowel feature dissimilarity ratio, indicating that higher phonetic distance and phonological dissimilarity between stimulus and response vowels predicts lower rate of perceptual confusions. Lastly, there was a significant and negative main effect of PMI ($\beta = -0.22$, $SE = 0.06$, $z = -3.68$, $p < 0.001$), indicating that, contrary to my hypothesis, *higher* likelihood of vowel cooccurrence between two vowels leads to *fewer* perceptual confusions (Figure 4.5).

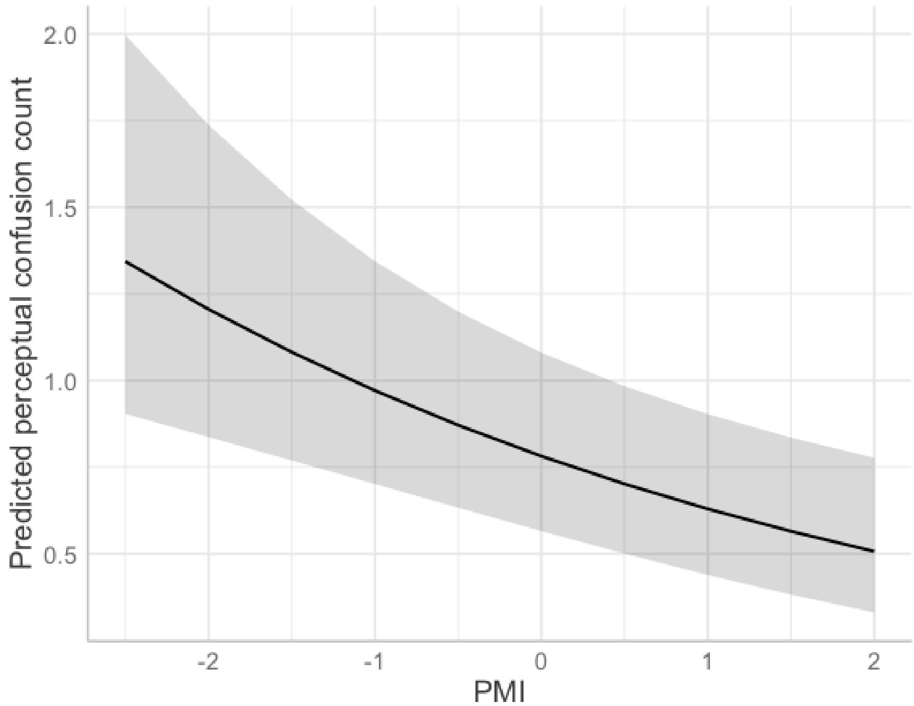

Figure 4.5. Predicted perceptual confusion count as a function of PMI.

4.3. Interim discussion

I modeled perceptual confusions in Turkish vowel identification as a measure of perceived vowel similarity with phonetic and phonological predictors to examine their relative contributions. Among the phonetic and phonological similarity measures, Euclidean distance had the largest effect on perceptual confusion rates whereas PMI had the smallest effect. These results suggest that perceived vowel similarity in Turkish is mainly predicted by the phonetic distance between a stimulus and a vowel category's centroid. Hall and Hume (2015) similarly found that Euclidean distance was the strongest predictor of perceived vowel similarity in French vowel identification. Together, these findings suggest that perceived vowel similarity is largely determined by vowel phonetics in both Turkish and French. The significant effects of Euclidean distance, Pillai score, and vowel feature dissimilarity ratio are all in the expected direction, whereby fewer perceptual confusions are predicted for vowels with greater phonetic distance or phonological dissimilarity. This described effect of vowel feature dissimilarity ratio is in line with Terbeek's (1977) finding that dissimilarity in backness leads to lower perceived similarity for Turkish listeners and extends this finding to the other Turkish vowel features. In contrast, the significant effect of PMI suggests that vowels with *higher* likelihood of cooccurrence in the lexicon are predicted to exhibit *fewer* perceptual confusions. In other words, I find that contrary to my hypothesis, higher phonotactic similarity leads to *lower* perceived similarity.

I suggest that it might be insightful to interpret the unexpected direction of the main effect of PMI in relation to the observations about vowel peripherality and Turkish vowel harmony summarized above (Clements & Sezer, 1982; Suomi, 1983). Recall that Clements and Sezer and Suomi had three claims regarding vowel peripherality in Turkish. The first claim is that nonperipheral Turkish vowels /ɯ œ y/ are perceptually compromised relative to the phonetically more prominent peripheral vowels (Suomi, 1983). My Turkish vowel identification accuracy analysis in the preceding section supported this first claim, showing that, in line with previous work, more peripheral vowels were identified with higher accuracy (e.g., Hall & Hume, 2015; Hillenbrand et al., 1995; Peterson & Barney, 1952). The second claim is that nonperipheral Turkish vowels are subject to a greater degree of vowel cooccurrence constraints (Clements & Sezer, 1982). I return to evidence supporting this claim later in this chapter.

The third claim is that Turkish vowel harmony facilitates the accurate identification of nonperipheral Turkish vowels (Suomi, 1983). To be exact, Suomi claims that nonperipheral Turkish vowels are identified with higher accuracy in harmonic contexts than in disharmonic contexts, as follows. Suomi argues that nonperipheral vowels are perceptually confusable with vowels with more extreme F2 values on the same height level. However, Suomi also observes that Turkish vowel harmony restricts vowel cooccurrence to vowels on the same F2 level as defined by backness and rounding. Consequently, nonperipheral vowels are not permitted to cooccur with vowels on their height level that have more extreme F2 values, i.e., the vowels that they are perceptually confused with. Stated in terms of vowel cooccurrence in the lexicon, Suomi's claim is that higher perceptual similarity drives lower likelihood of vowel cooccurrence in the lexicon, and that this effect is stronger for nonperipheral Turkish vowels. In other words, Suomi argues that there is a negative relationship between perceptual similarity of vowels and their likelihood to cooccur. Note that in Suomi’s perspective, the relationship is one in which perceived similarity predicts likelihood of vowel cooccurrence. In contrast, I explore how likelihood of vowel cooccurrence might predict perceived vowel similarity, and I hypothesized a positive relationship. Nonetheless, stated in terms of my study, Suomi’s claim calls for the reverse of my initial hypothesis. The new hypothesis is that *higher* likelihood of vowel cooccurrence might be an indicator of *lower* perceived vowel similarity.

To reiterate, Suomi (1983) and Clements and Sezer (1982) claim that perceptual similarity might be a predictor of likelihood of vowel cooccurrence in Turkish with higher perceptual similarity leading to a higher degree of cooccurrence constraints and hence lower likelihood of cooccurrence. A prediction that follows from applying this line of reasoning to the present study is that *higher* likelihood of vowel cooccurrence might predict *fewer* perceptual confusions in Turkish vowel identification, and that this effect might be driven by nonperipheral Turkish vowels. To test this new hypothesis, I present a post-hoc zero-inflated Poisson regression analysis where I add stimulus vowel peripherality and the interaction of stimulus vowel peripherality and PMI as predictors.

4.4. Poisson regression with vowel peripherality as a predictor

Although the effects of Euclidean distance, Pillai score, and vowel feature dissimilarity ratio were in the expected direction in the zero-inflated Poisson regression reported above, the direction of the effect of PMI was contrary to my hypothesis of higher likelihood of cooccurrence leading to more perceptual confusions. To explore whether this effect was driven by nonperipheral vowels, I ran a post-hoc zero-inflated Poisson regression with stimulus vowel peripherality (sum-coded with nonperipheral vowel as -1 and peripheral vowel as 1) and the interaction of stimulus vowel peripherality and PMI as additional predictors. I predict a significant main effect of stimulus vowel peripherality whereby higher rates of perceptual confusions are predicted for nonperipheral stimulus vowels compared to peripheral stimulus vowels. I also predict a significant interaction of stimulus vowel peripherality and PMI such that higher likelihood of vowel cooccurrence predicts fewer perceptual confusions for nonperipheral stimulus vowels than for peripheral stimulus vowels. A likelihood ratio test comparing the post-hoc model with the original model revealed that the model with the additional parameters significantly improved model fit ($\chi^2(4) = 95.46$, $p < 0.001$). The results from the logistic regression part of the model are reported in Table 4.5 and the results from the Poisson regression part of the model are reported in Table 4.6.

Table 4.5. Model results from the post-hoc binomial logistic regression.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.46	0.33	4.45	< 0.001***
F1	-0.32	0.19	-1.69	0.09
F2	-0.71	0.28	-2.56	0.01*
Speaker gender	0.42	0.26	1.61	0.11
Duration	0.02	0.15	0.13	0.89
Euclidean distance	1.56	0.33	4.69	< 0.001***
Pillai score	1.52	0.43	3.55	< 0.001***
Feature dissimilarity ratio	0.91	0.22	4.13	< 0.001***
PMI	0.49	0.17	2.82	0.005**
Stimulus vowel peripherality	0.88	0.31	2.79	0.005**
F2 : Speaker gender	-0.13	0.32	-0.42	0.68
PMI : stimulus vowel peripherality	0.24	0.40	0.61	0.54

As in the previous model's binomial logistic regression, there were significant main effects of stimulus vowel F2, Euclidean distance, Pillai score, vowel feature dissimilarity ratio, and PMI, all of which were of similar magnitude and were in the same direction in both models. The additional predictor vowel peripherality also had a significant and positive main effect, indicating that the model predicts no perceptual confusions at all with higher likelihood when the stimulus vowel is a peripheral vowel.

Table 4.6. Model results from the post-hoc Poisson regression.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-0.12	0.17	-0.68	0.50
F1	0.02	0.10	0.25	0.81
F2	0.55	0.13	4.20	< 0.001***
Speaker gender	0.13	0.10	1.33	0.18
Duration	-0.32	0.06	-5.45	< 0.001***
Euclidean distance	-1.45	0.16	-8.86	< 0.001***
Pillai score	0.04	0.12	0.33	0.74
Feature dissimilarity ratio	-0.42	0.07	-5.93	< 0.001***
PMI	-0.39	0.07	-5.90	< 0.001***
Stimulus vowel peripherality	-0.30	0.13	-2.35	0.02*
F2 : Speaker gender	-0.24	0.14	-1.68	0.09
PMI : stimulus vowel peripherality	0.61	0.10	6.15	< 0.001***

The post-hoc Poisson regression model results paralleled the results in the original model, with significant main effects of stimulus vowel F2 and duration, Euclidean distance, and feature dissimilarity ratio that were in the same direction and of similar magnitude in both models. The significant main effects of vowel feature dissimilarity ratio and PMI became stronger in the post-hoc model, suggesting that higher vowel feature dissimilarity and higher likelihood of vowel cooccurrence predict lower rate of perceptual confusions. There was a significant and negative main effect of vowel peripherality, indicating that peripheral stimulus vowels had lower rates of perceptual confusion than nonperipheral stimulus vowels. Lastly, there was a significant interaction of vowel peripherality and PMI, indicating that whereas for nonperipheral stimulus vowels, higher likelihood of cooccurrence with a vowel predicted *fewer* perceptual confusions, for peripheral stimulus vowels, higher likelihood of cooccurrence with a vowel predicted *more* perceptual confusions (Figure 4.6).

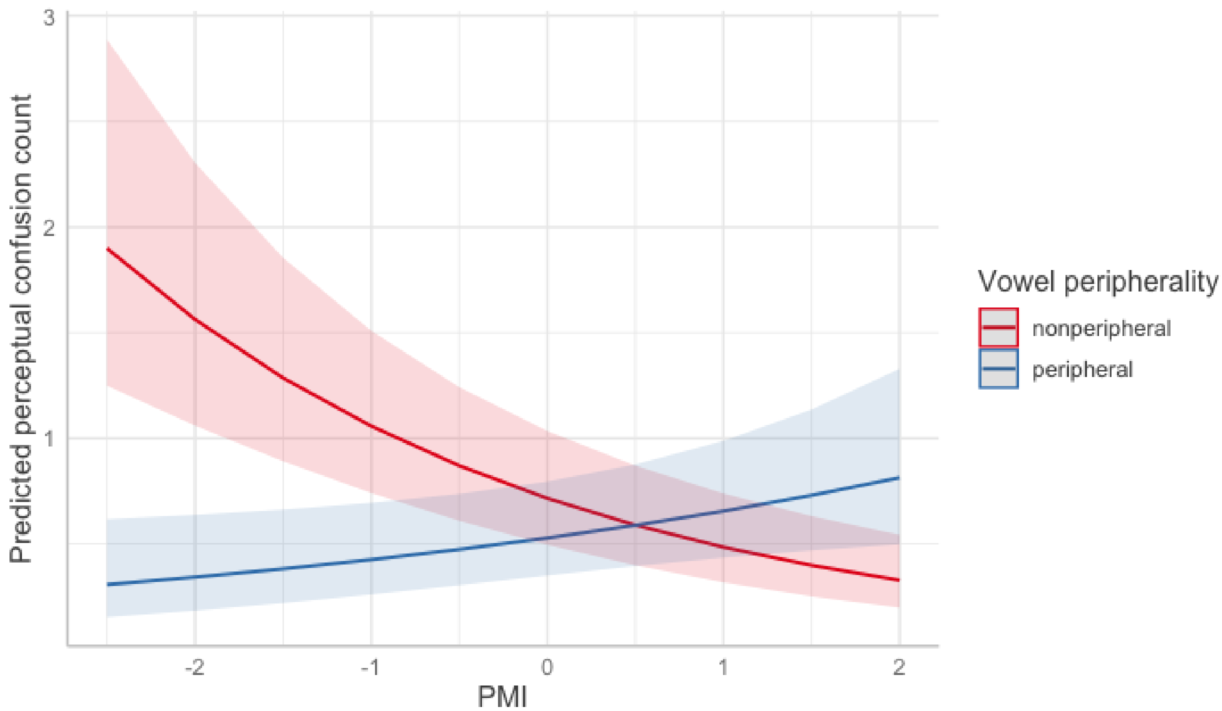

Figure 4.6. Predicted perceptual confusion count as a function of PMI and vowel peripherality.

This post-hoc analysis was run to explore whether the negative relationship between PMI and perceptual confusion rate indicating that higher likelihood of vowel cooccurrence predicts fewer perceptual confusions is driven by nonperipheral vowels. The results here are aligned with my predictions regarding stimulus vowel peripherality as an additional predictor. The main effect of stimulus vowel peripherality corroborates the findings from the vowel identification accuracy analysis, suggesting that nonperipheral stimulus vowels are more likely to be perceptually confused than peripheral stimulus vowels. The significant interaction of PMI and stimulus vowel peripherality indicate that the effect of PMI is qualified by stimulus vowel peripherality. Recall that I hypothesized that higher likelihood of vowel cooccurrence would predict more perceptual confusions, yet the main effect of PMI suggested that higher likelihood of vowel cooccurrence predicts *fewer* perceptual confusions. With my revised hypothesis, following Suomi (1983), this unexpected effect of PMI is enhanced for nonperipheral stimulus vowels, suggesting that the effect is driven by nonperipheral stimulus vowels. However, for peripheral stimulus vowels, the interaction suggests that higher likelihood of vowel cooccurrence predicts *more* perceptual confusions. Hence, I find evidence supporting my original hypothesis that phonotactic similarity is positively related to perceived vowel similarity, although this effect is relatively weak, is limited to peripheral vowels, and does not generalize to all vowels as I had originally hypothesized. I elaborate on the implications of this pattern of results in the general discussion.

One remaining question is whether the significant relationship between likelihood of vowel cooccurrence and perceived vowel similarity is a consequence of vowel harmony or generalizes to a nonharmony language. To this end, I present a comparative analysis of English vowel identification data from Hillenbrand et al. (1995) in the following section. I hypothesize that likelihood of vowel cooccurrence in the lexicon will not be a significant predictor of perceptual confusions in English as vowel cooccurrence is relatively unconstrained and hence arbitrarily related to vowel identity in English unlike in Turkish.

4.5. English vowel identification

In their study aiming to describe the vowel acoustics of American English, Hillenbrand et al. (1995) collected 12 American English vowels produced in /hVd/ contexts from 139 speakers of the ‘upper Midwest’ dialect of American English. A total of 1668 /hVd/ productions were presented to a different group of 20 ‘upper Midwest’ American English listeners in a 12-alternative forced choice (12AFC) vowel identification experiment, where listeners had to identify the syllable they heard among 12 English words such as *heed*, *hid*, *head*, etc. Hillenbrand et al. (1995) has publicly available count data by each of the stimulus tokens, where a total of 22,172 responses across all participants are reported as the number of times a stimulus token is identified as each of the 12 response vowel alternatives (represented in Figure 4.7 as a heatmap of response proportions). Listeners accurately identified 95.29% of the vowels in the stimuli they heard, and there were 1042 perceptual confusions in total. Note that in the 8AFC Turkish vowel identification experiment, accuracy was considerably lower at 89.53%. As no by-participant response information is provided by Hillenbrand et al. (1995), the data cannot be modeled with by-participant random effects. Moreover, multinomial logistic regressions with random intercepts for item did not converge. As the data could not be modeled with random effects structure, I only report a Poisson regression analysis of the vowel identification count data.

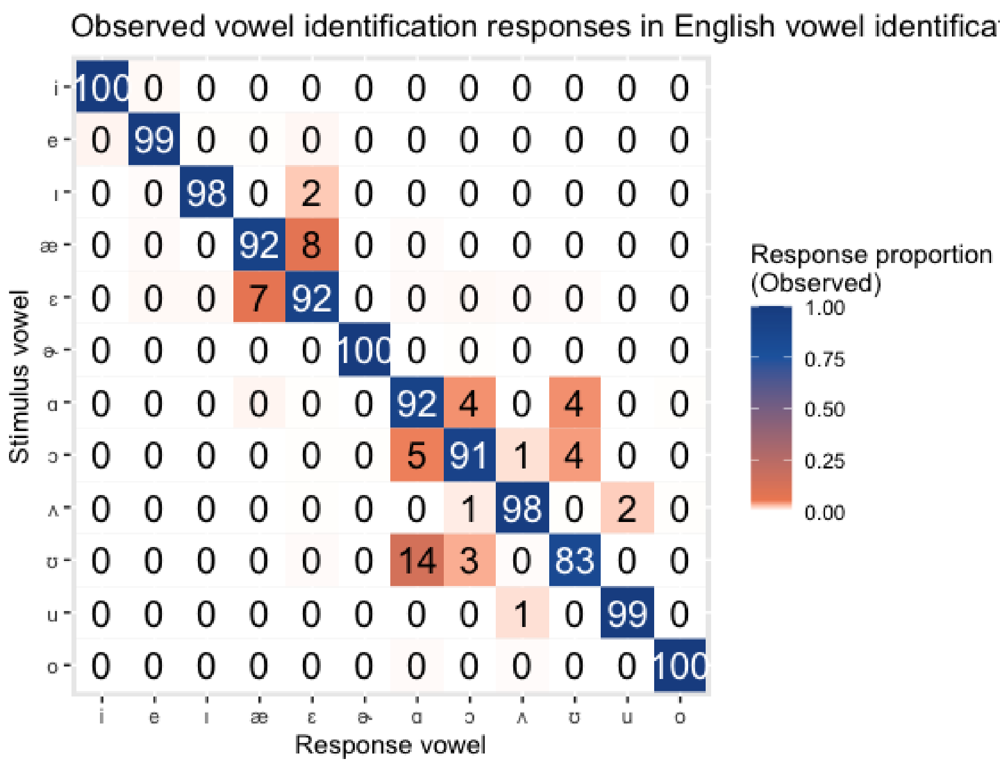

Figure 4.7. Heatmap depicting the observed response proportions in the Hillenbrand et al. (1995) English vowel identification task.

4.5.1. Results and discussion

I modeled the Hillenbrand et al. (1995) vowel identification count data in a zero-inflated Poisson regression (`pscl` package; Jackman, 2024; Zeileis et al., 2008) to simultaneously account for systematic zero count responses and model variation in response counts. Following the zero-inflated Poisson regression models for the Turkish vowel identification count data, the predictors in the model were stimulus vowel F1 and F2, speaker gender, the interactions of stimulus vowel F1 and F2 with speaker gender, Euclidean distance between stimulus vowel and response vowel category centroid, Pillai score between stimulus vowel category and response vowel category, vowel feature dissimilarity ratio between stimulus and response vowel categories, and PMI between stimulus and response vowel categories. Each phonetic predictor was calculated over the phonetic values provided for the English vowel stimuli in Hillenbrand et al. (1995). Vowel feature dissimilarity was calculated over 4 features (backness, rounding, height, and tenseness), and PMI was calculated over 61,905 disyllabic English words in the CMU Pronouncing Dictionary. The same set of predictors were included in both the logistic regression and the Poisson regression parts of the model, with the logistic regression modeling the 0 counts and the Poisson regression modeling the variation in the count data. Recall that I predict Euclidean distance, Pillai score, and vowel feature dissimilarity ratio to have significant and positive main effects in the logistic regression and significant and negative main effects in the Poisson regression. As vowel cooccurrence is relatively unconstrained in English, I predict PMI to have a weak effect, and to have a negative effect in the logistic regression and a positive effect in the Poisson regression. I report the results from the logistic regression in Table 4.7 and the results from the Poisson regression in Table 4.8.

Table 4.7. Model results from the logistic regression modeling the 0 counts in the English vowel identification data.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	3.95	0.16	24.85	< 0.0001***
Vowel F1	-0.18	0.07	-2.70	0.01**
Vowel F2	0.37	0.09	4.27	< 0.0001***
Speaker gender	-0.15	0.07	-2.16	0.03*
Vowel duration	0.07	0.05	1.45	0.15
Euclidean distance	-0.13	0.05	-2.57	0.01*
Pillai score	1.22	0.07	16.45	< 0.0001***
Vowel feature dissimilarity ratio	-1.41	0.26	-5.43	< 0.0001***
PMI	0.27	0.06	4.73	< 0.0001***
Speaker gender : Vowel F1	-0.01	0.06	-0.22	0.83
Speaker gender : Vowel F2	-0.25	0.08	-3.27	0.0001***

The results from the logistic regression suggest that stimulus vowel F1 and F2 had significant main effects, indicating that the model predicts no perceptual confusions at all with a higher likelihood when the stimulus vowel has lower F1 and higher F2. There were significant yet negative main effects of Euclidean distance and vowel feature dissimilarity ratio, indicating that as phonetic and phonological dissimilarity between stimulus-response vowels increase, the likelihood of observing no perceptual confusions at all decreases. There was a significant and positive main effect of Pillai score, suggesting that as phonetic overlap between vowel categories decrease, the likelihood of observing no perceptual confusions at all increases. There was a significant interaction of speaker gender and vowel F2, indicating that stimulus vowels with higher F2 produced by female speakers are less likely to yield no perceptual confusions at all compared to male speakers. Lastly, there was a significant and positive main effect of PMI, suggesting that as the likelihood of two vowels to cooccur increases, the likelihood of observing no perceptual confusions at all also increases, indicating that vowels that are more likely to cooccur are also more likely to be not confused.

Table 4.8. Model results from the Poisson regression modeling the English vowel identification count data.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.37	0.13	2.85	0.004**
Vowel F1	0.35	0.05	6.30	< 0.0001***
Vowel F2	-0.07	0.06	-1.16	0.24
Speaker gender	-0.07	0.05	-1.41	0.16
Vowel duration	-0.01	0.04	-0.30	0.77
Euclidean distance	0.17	0.03	5.75	< 0.0001***
Pillai score	-0.76	0.07	-10.83	< 0.0001***
Vowel feature dissimilarity ratio	-0.75	0.22	-3.42	< 0.0001***
PMI	0.27	0.05	5.24	< 0.0001***
Speaker gender : Vowel F1	-0.19	0.05	-3.68	< 0.0001***
Speaker gender : Vowel F2	-0.27	0.05	-5.44	< 0.0001***

The results from the Poisson regression were generally in line with the results from the logistic regression. Stimulus vowel F1 had a significant main effect indicating that as stimulus vowel F1 increases, perceptual confusion rate increases. The significant yet positive main effect of Euclidean distance indicates that as the phonetic distance between a stimulus vowel and a response vowel category increases, contrary to the predicted effect of phonetic dissimilarity, the perceptual confusion rate of these vowels also increases. There were significant and negative main effects of Pillai score and vowel feature dissimilarity ratio, indicating that stimulus and response vowel categories that have lesser phonetic overlap and fewer shared vowel features have a lower perceptual confusion rate. There were significant interactions of speaker gender and stimulus vowel F1 and F2, indicating that as stimulus vowel F1 and F2 increase, stimulus vowels have lower perceptual confusion rates if they are produced by female speakers. Lastly, there was a significant and positive main effect of PMI ($\beta = 0.27$, $SE = 0.05$, $z = 5.24$, $p < 0.0001$), indicating that higher likelihood of vowel cooccurrence predicts more perceptual confusions.

Recall that in the French (Hall & Hume, 2015) and Turkish vowel identification experiments, Euclidean distance was the strongest predictor of perceived vowel similarity with greater distance predicting more perceptual confusions. In the English vowel identification data, although Euclidean distance is one of the significant predictors, its effect is unexpectedly in the opposite direction. In their own analysis of the data, Hillenbrand et al. (1995) ran discriminant analyses predicting vowel categories from phonetic descriptors of the stimuli. Their comparison of various predictors revealed that vowel formants measured at 2 timepoints rather than a single timepoint improved vowel classification accuracy. Following their discussion, I speculate that Euclidean distance measured at vowel midpoint might be similarly misleading as a predictor of the vowel identification responses and not be representative of the spectral changes that influence English vowel identification to a greater degree. In contrast to the French and Turkish results, the strongest predictor of perceived vowel similarity in English is Pillai score such that greater phonetic overlap between vowel categories predict more perceptual confusions. Although the unexpected direction of Euclidean distance is unexplained, the finding that Pillai score is the strongest predictor of perceived vowel similarity in English is also in line with the results from the English vowel discrimination experiments reported in Chapter 3.

As in the Turkish vowel identification experiment, the main predictor of interest was PMI: I predicted that PMI would be either a stronger predictor of perceived vowel similarity in Turkish than in English or not be a significant predictor in English due to vowel cooccurrence in the English lexicon not being constrained by vowel harmony. That PMI is a significant predictor of perceptual confusion rates in English vowel identification suggests that the likelihood of vowel cooccurrence influences vowel identification not only in a harmony language but also in a nonharmony language. Moreover, that PMI has a positive relationship with perceptual confusion rates is in line with my main hypothesis in this work that higher likelihood of vowel cooccurrence predicts higher rate of perceptual confusions. This finding suggests that vowels that are more likely to cooccur and hence phonotactically more similar are perceived to be more similar even in a language where vowel cooccurrence is mostly unconstrained. However, this finding stands in contrast with the findings from the Turkish vowel identification experiment.

In the Poisson regression for the Turkish vowel identification data with the same predictors, PMI had a significant yet negative main effect (see Table 4.4), suggesting that higher likelihood of vowel cooccurrence leads to *lower* perceived similarity. In the post-hoc Poisson regression modeling the Turkish vowel identification data, PMI had a significant interaction with stimulus vowel peripherality whereby higher likelihood of vowel cooccurrence led to higher perceived similarity for peripheral stimulus vowels only (see Table 4.6). That higher likelihood of vowel cooccurrence predicts higher perceived similarity for all vowels in English but not in Turkish is unexpected and suggests that the likelihood of vowel cooccurrence plays different roles in Turkish and English vowel perception. The question of whether the effect is specific to harmony languages or stronger in harmony languages than in nonharmony languages is complicated by the opposite effects of likelihood of vowel cooccurrence in the two languages and remains unanswered.

4.6. General discussion

In this chapter, I presented results from a Turkish vowel identification experiment where I analyzed perceptual confusions as a measure of perceived vowel similarity with phonetic, phonological, and phonotactic measures of vowel similarity as predictors. Of particular interest was the likelihood of vowel cooccurrence in the lexicon as a phonological-phonotactic measure of vowel similarity and a predictor of perceived vowel similarity. I hypothesized that vowels that are more likely to cooccur in Turkish would be coactivated more and hence perceived to be more similar. I find that the likelihood of vowel cooccurrence is a significant predictor of vowel identification in Turkish and yet, contrary to my hypothesis, a higher likelihood of vowel cooccurrence predicts a lower rate of perceptual confusions in Turkish. I discuss this finding in relation to vowel peripherality as an additional predictor of perceived vowel similarity.

A post-hoc analysis revealed that higher likelihood of vowel cooccurrence predicts lower rate of perceptual confusions for nonperipheral Turkish vowels but higher rate of perceptual confusions for peripheral Turkish vowels. In other words, higher likelihood of vowel cooccurrence leads to higher perceived similarity, but only when the stimulus vowel is peripheral. The opposite is true for nonperipheral stimulus vowels, suggesting that phonotactic similarity might play different roles for peripheral and nonperipheral Turkish vowels.

4.6.1. Vowel cooccurrence and harmony classes in Turkish

My main research question in this work is about whether harmony classes and/or vowel features in Turkish emerge from vowel cooccurrence patterns in the Turkish lexicon, which I examined by testing likelihood of vowel cooccurrence as a predictor of perceived confusability of vowels. Recall that this research question stems from the exemplar models view that abstract phonological representations are similarity-based generalizations over exemplars. In other words, the question is about identifying what constitutes the similarity of exemplars for harmony classes to emerge. Phoneme categories emerge from the coactivation of exemplars, i.e., phonetically specified vowel tokens, based on their phonetic similarity. In comparison, a harmony class is an abstraction over more than one phoneme category, and presumably involves the coactivation of exemplars beyond what is predicted by their phonetic similarity alone. Based on previous computational work suggesting that vowel cooccurrence patterns in the Turkish lexicon predict harmony classes (e.g., Baker, 2009; Cole, 2009), I hypothesized that higher likelihood of vowel cooccurrence in the lexicon would lead to higher perceived vowel similarity. Under this hypothesis, likelihood of vowel cooccurrence is instrumental in enhancing the similarity of vowels beyond their phonetic similarity, thereby leading to emergent harmony classes.

In order to explore the factors that influence perceived vowel similarity, I modeled perceptual confusions in Turkish vowel identification with phonetic, phonological, and phonotactic measures of vowel similarity. Table 4.9 presents a summary of the significant predictors. In line with previous research (e.g., Hall & Hume, 2015), the results here suggest that perceived vowel similarity in Turkish is mainly predicted by the phonetic similarity of exemplars. The strongest predictor of perceptual confusion rates is Euclidean distance, with a main effect suggesting that smaller phonetic distance between the stimulus vowel token and the response vowel category's centroid leads to higher perceived similarity. The significant main effect of vowel feature dissimilarity ratio suggests that vowels with a higher number of shared features leads to higher perceived similarity. This finding is also in line with the insight from Terbeek (1977), who found that Turkish listeners perceived nonnative front-back vowel pairs to be less similar than listeners of other languages, suggesting that perceived vowel similarity in Turkish is influenced by the phonological similarity of vowels in terms of their distinctive features. The results here extend this insight and suggests that when Turkish listeners process Turkish vowels on a phonological level rather than on a purely phonetic level, greater phonological similarity in vowel features enhances the perceived similarity of vowels beyond their phonetic similarity.

Table 4.9. Summary of significant results of Poisson regression modeling perceptual confusions in Turkish vowel identification.

Significant predictor	Explanation	Predictor type	Direction of the effect
Euclidean distance	Measured between the stimulus vowel token at vowel midpoint and the response vowel category's centroid	Phonetic similarity	Higher similarity leads to higher perceived similarity
Vowel feature dissimilarity ratio	Measured between the stimulus-response vowel categories	Phonological similarity	Higher similarity leads to higher perceived similarity
PMI	Stimulus-response vowel pair's likelihood of cooccurrence in the lexicon	Phonotactic similarity	Higher phonotactic similarity leads to lower perceived similarity

In contrast to the effect of phonological similarity, and contrary to my hypothesis, I find that higher likelihood of vowel cooccurrence (PMI) in the Turkish lexicon predicts fewer perceptual confusions. In other words, higher phonotactic similarity leads to *lower* perceived vowel similarity, suggesting that vowels that are phonotactically more similar and more likely to be coactivated by a lexical item are perceptually less similar. This finding is unexpected based on the general prediction that greater phonetic, phonological, or phonotactic similarity of vowels would be reflected in greater perceived similarity. This finding suggests that unlike greater phonological similarity in vowel features, greater phonotactic similarity in vowel cooccurrence does not enhance but rather lessens perceived vowel similarity. As such, the findings here do not support the view that harmony classes in Turkish emerge from vowel cooccurrence patterns in the lexicon. Moreover, in a harmony language such as Turkish, vowels that share a higher number of features are expected to be highly likely to cooccur in the lexicon. Hence, these phonological and phonotactic similarity measures are predicted to have effects in the same direction. That the effects of vowel feature dissimilarity ratio and PMI are in opposite directions suggests that phonotactic similarity might play a different role in Turkish than what I hypothesized. I discuss one alternative role in the next section.

Lastly, I modeled a dataset of English vowel identification (Hillenbrand et al., 1995) with the same set of predictors for a cross-linguistic comparison with Turkish. I hypothesized that higher phonotactic similarity would lead to higher perceived vowel similarity in both English and Turkish, and yet this effect would be weaker in English compared to Turkish as there are no harmony classes in English that can be defined by vowel cooccurrence patterns in the English lexicon. The results reveal that PMI is a significant predictor such that vowels that are more likely to cooccur in the English lexicon had higher perceptual confusion rates in English vowel identification. Hence, the significant main effect of PMI in English suggests that phonotactic similarity might have a role in enhancing perceived vowel similarity as I originally hypothesized. This original hypothesis linked phonotactic similarity (as measured by PMI here) to perceived vowel similarity in Turkish as well as in English and argued that this link was instrumental in emergent harmony classes in Turkish. As there are no harmony classes to be explained and linked to vowel phonotactics in English, this particular result does not address whether phonotactic similarity leads to harmony classes to emerge. In other words, although the results here provide supporting evidence that higher phonotactic similarity is associated with higher perceived vowel similarity as observed in English, I do not find evidence supporting that this positive association is what underlies harmony classes in Turkish.

4.6.2. Vowel peripherality and vowel cooccurrence in Turkish

To better understand the unexpected effect of PMI in Turkish vowel identification whereby higher phonotactic similarity predicts lower perceived vowel similarity, I ran an exploratory, post-hoc Poisson regression with stimulus vowel peripherality as an additional predictor. This post-hoc analysis was motivated by previous work suggesting that Turkish vowel harmony treats peripheral and nonperipheral vowels differently based on their phonetic and perceptual differences: Clements and Sezer (1982) and Suomi (1983) argue that the nonperipheral Turkish vowels are perceptually more confusable than the phonetically more extreme peripheral vowels, and are subject to a greater degree of vowel cooccurrence restrictions to facilitate their accurate identification via contextual cues. Following these claims that there is a stronger relationship between perceptual confusability and phonotactic restrictions for the nonperipheral vowels, I hypothesized that the nonperipheral vowels would drive the unexpected effect of PMI.

The post-hoc Poisson regression modeling the Turkish perceptual confusions confirmed that vowel peripherality is one of the significant predictors of perceptual confusions in Turkish. The post-hoc model results revealed that peripheral Turkish vowels have lower rates of perceptual confusion, which replicates the finding that vowel identification accuracy is higher for peripheral Turkish vowels. In addition, there is a significant interaction of vowel peripherality and likelihood of vowel cooccurrence. This significant interaction suggests that likelihood of vowel cooccurrence plays different roles in the perception of peripheral and nonperipheral vowels.

For peripheral Turkish vowels, the significant interaction yields results that are in line with my main hypothesis that higher likelihood of vowel cooccurrence fosters higher perceived vowel similarity. This findings supports the interpretation that likelihood of vowel cooccurrence as a measure of phonotactic similarity is one of the predictors of perceived vowel similarity. Under this interpretation, phonotactic similarity enhances the perceived similarity of vowels beyond their phonetic similarity and hence might lead to the emergence of harmony classes. However, this interpretation only applies to peripheral Turkish vowels as the effect of phonotactic similarity on perceived similarity does not generalize to all Turkish vowels in the form of a main effect. Hence, this finding does not support my hypothesis that Turkish harmonic vowel classes, which include both peripheral and nonperipheral vowels, emerge from vowel cooccurrence patterns. In fact, the significant interaction shows that the main effect in the unexpected direction is driven by nonperipheral vowels: higher phonotactic similarity leads to lower perceived vowel similarity for nonperipheral vowels and this effect is strong enough to dominate over the effect in the opposite direction for peripheral vowels as well.

Together, these findings suggest that phonotactic similarity does not in a general manner enhance perceived vowel similarity among vowels within a harmony class, as I predicted. Hence, why phonotactic similarity might have an effect in the opposite direction requires an explanation. My post-hoc analysis with vowel peripherality as a predictor was inspired by Suomi (1983). Hence, for a possible explanation I turn to Suomi's main idea that Turkish vowel harmony is perceptually motivated to prevent nonperipheral Turkish vowels from being perceptually confused. Suomi claims that, especially when nonperipheral vowels with perceptually less prominent contrasts are concerned, restricting vowel cooccurrence *contextually* facilitates predicting and accurately identifying vowels. Similarly, I interpret the findings here to suggest that phonotactic similarity enhances the perceptual *dis*similarity of Turkish vowels rather than their similarity and hence might have a role in preserving and enhancing weaker phonemic contrasts perceptually.

The findings here reflect perceptual confusions in vowel identification where the vowel stimuli are presented in monosyllabic nonwords. In daily speech, Turkish words are not restricted to monosyllabic words. Güngör (2003) analyzes a Turkish corpus of 2 million words scraped from newspapers, periodicals, and novels, and finds that average word length in this corpus is 6.13 characters, which indicates that the average Turkish word is disyllabic based on the fact that maximal syllable structure in Turkish is CCVCC and consonants are not syllabic. Thus, when a nonperipheral Turkish vowel occurs in the average Turkish word, it occurs in the context of another vowel. This contextual vowel's phonemic identity is predicted by a measure of likelihood of vowel cooccurrence such as PMI calculated over the disyllabic words in the Turkish lexicon as in the present work. The significant interaction of vowel peripherality and likelihood of vowel cooccurrence in the post-hoc analysis suggests that, in the average disyllabic Turkish word, the nonperipheral vowels cooccur with the vowels with which they are less likely to be perceptually confused. These observations suggest that, as proposed by Suomi (1983), Turkish vowel harmony might help the accurate perception of the nonperipheral vowels by limiting their cooccurrence with the vowels with which they are confusable. In other words, by positionally and contextually making nonperipheral vowels as predictable, vowel harmony might reduce perceptual confusions for these vowels.

To more directly test the claim that the stronger cooccurrence restrictions on nonperipheral Turkish vowels are perceptually motivated to facilitate their accurate identification (Clements & Sezer, 1982; Suomi, 1983), I ran a post-hoc linear regression modeling PMI as a function of stimulus vowel peripherality, perceptual confusion count in the Turkish vowel identification experiment, and their interaction (Table 4.10). There was a significant main effect of stimulus vowel peripherality, confirming that nonperipheral Turkish vowels have overall lower likelihood of vowel cooccurrence indicating stronger vowel cooccurrence restrictions compared to peripheral Turkish vowels. There was also a significant main effect of perceptual confusion suggesting that lower likelihood of cooccurrence in the Turkish lexicon is predicted for vowels that were more often perceptually confused in vowel identification. Lastly, there was a significant interaction suggesting that the effect whereby a higher rate of perceptual confusion predicts lower likelihood of vowel cooccurrence is stronger for nonperipheral Turkish vowels than peripheral Turkish vowels (Figure 4.8). Additional subset analyses confirmed that there is a strong negative relationship between perceptual confusion rates and likelihood of vowel cooccurrence in nonperipheral vowels ($\beta = -0.34$, $SE = 0.08$, $t = -4.34$, $p = 0.0001$). In contrast, in peripheral Turkish vowels, although higher rates of perceptual confusion predicted higher likelihood of vowel cooccurrence, this effect was marginally significant ($\beta = 0.19$, $SE = 0.10$, $t = 1.93$, $p = 0.054$). Together, these findings support the claim that Turkish vowel harmony might be motivated by a goal to reduce perceptual confusions by restricting vowel cooccurrence. To my knowledge, I show for the first time that perceived vowel similarity might predict the likelihood of vowel cooccurrence in the Turkish lexicon.

Table 4.10. Results of the post-hoc linear regression modeling PMI in the Turkish lexicon.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.84	0.09	-9.03	< 0.001***
Peripherality	1.00	0.12	8.47	< 0.001***
Perceptual confusion count	-0.34	0.06	-5.29	< 0.001***
Perceptual confusion count : Peripherality	0.54	0.14	3.93	< 0.001***

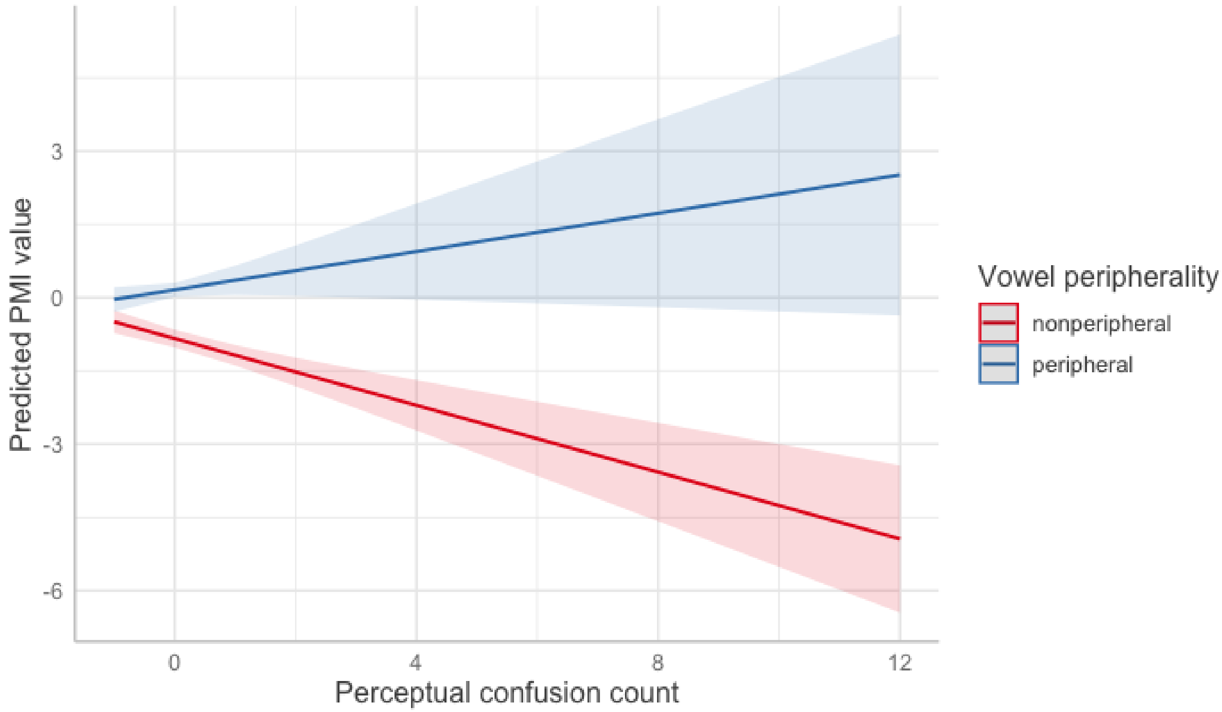

Figure 4.8. Likelihood of vowel cooccurrence in the Turkish lexicon as predicted by vowel peripherality and perceptual confusion count in Turkish vowel identification.

To reiterate, I propose with Suomi (1983) that Turkish vowel harmony might help vowel perception by limiting the cooccurrence of perceptually confusable nonperipheral vowels. One case that might go against this proposed relationship is the vowel pair /œ y/. Both of these vowels are nonperipheral and highly likely to be perceptually confused with each other, and yet have a high likelihood of cooccurrence. This pattern of high likelihood of cooccurrence and high rate of perceptual confusion is paralleled in the vowel pair /o u/, which are peripheral unlike /œ y/. Notice that /o œ u y/ are rounded vowels and both /œ y/ and /o u/ are backness harmonic vowel pairs. These rounded and backness harmonic vowel pairs have an additional constraint on their cooccurrence: both rule-based and lexically grounded descriptions of Turkish vowel harmony observe that the nonhigh rounded vowels /o œ/ are not attested or not allowed in non-initial syllables of Turkish words (e.g., Kabak, 2011; Kabak et al., 2008; Suomi, 1983). In my dataset of disyllabic Turkish words, the vowel sequence /u o/ is attested in only 49 words, and the vowel sequence /y œ/ is attested in only 7 words. In contrast, the vowel sequence /o u/ is attested in 437 words and the vowel sequence /œ y/ is attested in 256 words. Hence, the rounded vowel pairs /u o/ and /y œ/ that are highly likely to be perceptually confused can be positionally disambiguated in the lexicon, supporting the proposal that vowel cooccurrence constraints in Turkish are perceptually motivated.

4.7. Conclusions

To summarize, this chapter presented a vowel identification study in Turkish aiming to discover (1) whether likelihood of vowel cooccurrence influences perceived vowel similarity beyond phonetic similarity and (2) whether harmony classes in Turkish emerge as a generalization over vowel exemplars that are coactivated based on phonotactic and perceptual similarity. I found mixed results with respect to the first question, as the likelihood of vowel cooccurrence in the Turkish lexicon and the rate of perceptual confusions in vowel identification had a significant relationship, yet higher phonotactic similarity leads to lower perceived vowel similarity in Turkish, which is not consistent with the hypothesis that harmony classes emerge from vowel cooccurrence patterns in the Turkish lexicon. However, a post-hoc analysis revealed that this effect is driven by nonperipheral Turkish vowels whereas phonotactic similarity enhances perceived similarity for peripheral vowels, as hypothesized. Despite the fact that this effect is in line with my hypothesis, it does not generalize to all Turkish vowels, which means that harmony classes in Turkish cannot be fully explained as emergent from vowel cooccurrence patterns in the lexicon by enhanced perceptual similarity between cooccurring vowels.

My hypothesis that Turkish harmony classes emerge from vowel cooccurrence patterns in the Turkish lexicon is grounded in the exemplar models framework. In exemplar models, a phonological category such as a harmony class emerges over a set of phoneme categories, whereas a phoneme category emerges over a set of exemplars based on their phonetic similarity. Based on computational work suggesting that phonotactic similarity of vowels lead to harmony classes to emerge (Baker, 2009; Cole, 2009), I hypothesized that phonotactic similarity would be reflected in enhanced perceived vowel similarity beyond the phonetic similarity of vowels. However, that nonperipheral Turkish vowels are inherently susceptible to perceptual confusions as evidenced by their low rates of accuracy in vowel identification led me to reconsider my original hypothesis. The goal of speech perception is accurate identification of speech sounds from phonetic and contextual cues. Enhanced perceptual similarity might lead to the weakening of phonemic contrasts to the point of neutralization, which would conflict with this goal. My hypothesis that vowel harmony leads to greater phonotactic and perceived similarity for vowels that are more likely to cooccur would not be expected to apply to vowels that have suboptimal phonemic contrasts, such as the nonperipheral Turkish vowels. Instead, a mechanism that enhances these weak phonemic contrasts might better suit the goal of speech perception, and I argued that Turkish vowel harmony might be viewed as one such mechanism (cf. Clements & Sezer, 1982; Suomi, 1983).

To conclude, the findings here underscore the role of phonetic similarity in emergent representations in exemplar models, including phonemic representations as well as representations spanning a group of phonemes such as a harmony class. Although harmony classes might emerge from phonotactic similarity when vowels are abstracted from their phonetics (Baker, 2009; Cole, 2009), the results here suggest that this abstraction is not well explained in terms of the perceived similarity among vowels in a harmony class.

CHAPTER 5

CONCLUSIONS

This dissertation approached vowel harmony from the perspective of exemplar models, which argues that exemplars are coactivated based on vowels cooccurrence, which leads to emergent harmony classes. Based on the bidirectional relationship between coactivation and similarity as argued by exemplar models, I hypothesized that higher likelihood of vowel cooccurrence in the lexicon leads to higher perceived vowel similarity in Turkish, a vowel harmony language. I presented results from a series of Turkish vowel production (Chapter 2), discrimination (Chapter 3), and identification (Chapter 4) experiments to examine Turkish vowels and harmony classes in the Turkish phonetic space and vowel similarity as perceived by native Turkish listeners. I also presented results from English vowel discrimination experiments (Chapter 3) and an analysis of English vowel identification data (Chapter 4) for a cross-linguistic comparison. In each of the chapters, my goals were as follows:

In Chapter 2, I presented a Turkish vowel production experiment where I aimed to explore the Turkish vowel space phonetically in controlled contexts limiting variation and in contexts inducing various types of phonetic variation. My main findings were that Turkish vowels are continuously distributed along the F2 dimension despite F2 cueing both vowel backness and rounding, that vowels do not participate in most phonological processes as a harmony class, and that phonetic variation might be conditioned by two converging goals: contrast enhancement and contrast preservation.

In Chapter 3, I aimed to explore the relative contributions of phonetic and phonological factors to perceived vowel similarity in a series of vowel discrimination experiments. I varied the task demands to prioritize either acoustic-phonetic or phonological processing of vowels. The

Turkish vowel discrimination results suggested that phonological knowledge has unique contributions to perceived vowel similarity. Moreover, I found that vowel features are more salient in Turkish vowel perception relative to English vowel perception, which I attributed to Turkish listeners' experience with vowel harmony. However, the results did not reveal any influence of the likelihood of vowel cooccurrence on Turkish vowel discrimination. Following these findings, I presented a Turkish vowel identification study in Chapter 4, where I predicted phonemic vowel identification to engage a higher degree of phonological vowel processing and reveal greater influence of phonotactic knowledge on perceived vowel similarity, including the knowledge of the likelihood of vowel cooccurrence in the lexicon.

In Chapter 4, I aimed to explore my main hypothesis that higher likelihood of vowel cooccurrence leads to higher perceived vowel similarity. The results from my analysis of the English vowel identification data supported this hypothesis. However, I found that the likelihood of vowel cooccurrence calculated over the Turkish lexicon has an effect on perceived vowel similarity that is not in the hypothesized direction. I interpreted this unexpected result to suggest that the likelihood of vowel cooccurrence might play a role other than enhancing perceptual similarities of vowels in Turkish. I presented post-hoc analyses with vowel peripherality as additional predictors where I found a significant interaction between vowel peripherality and the likelihood of vowel cooccurrence.

The post-hoc hypothesis followed from previous work suggesting that nonperipheral Turkish vowels are highly perceptually confusable along the F2 dimension, and that Turkish vowel harmony is perceptually motivated to facilitate the accurate perception of these vowels (Suomi, 1983). My own results from the Turkish vowel production study in Chapter 2 were in line with the claim that nonperipheral vowels have weak phonetic contrast in F2. Accordingly, following Suomi, I hypothesized that the negative relationship between the likelihood of vowel cooccurrence in the Turkish lexicon and perceived vowel similarity to be driven by nonperipheral vowels. The post-hoc analysis results revealed a significant interaction whereby the hypothesized positive relationship between the likelihood of vowel cooccurrence and perceived vowel similarity holds for peripheral Turkish vowels. On the other hand, the interaction revealed a stronger negative relationship for the nonperipheral Turkish vowels. I argued that given their perceptual confusability, enhanced perceived similarity might lead to contrast neutralization for nonperipheral Turkish vowels, and that Turkish vowel harmony might be motivated to contextually enhance these weak phonetic contrasts by cueing vowel F2 redundantly across segments, as proposed by Suomi and Kaun (2004).

To reiterate, I find partial support for my main hypothesis that higher likelihood of vowel cooccurrence leads to higher perceived similarity. I argue that this proposed relationship is qualified by phonetic and perceptual confusability: for vowels that are sufficiently distinct in the phonetic vowel space, higher likelihood of vowel similarity enhances perceived similarity, which might lead to emergent harmony classes. On the other hand, for vowels that are not optimally distinct in the phonetic vowel space, vowel harmony might enhance contrasts. By limiting the perceptually confusable nonperipheral vowels to contexts in which they are predictable and cues to weak phonetic contrasts are extended across segments, Turkish vowel harmony might sustain the suboptimal vowel system as I will argue below.

Typologically, vowel systems with optimal phonetic dispersion and minimal perceptual confusability are preferred. For instance, McGahay (2024) shows through simulations that vowel systems with high confusability and low dispersion go through sound change to achieve optimal dispersion with minimal confusability. In these simulations, nonperipheral vowels that are attested in languages are predicted under the condition that they achieve minimal perceptual confusability and maximal distinction from other vowels in that vowel space. This simulated result is in line with the observation that typologically, schwa is the preferred nonperipheral vowel, which is phonetically maximally distinct from the peripheral vowels.

In Turkish, there are three nonperipheral vowels (/ɯ œ y/) that are not maximally distinct from neither the peripheral vowels nor the other nonperipheral vowels (see Chapter 2). Moreover, the Turkish vowel identification results I presented here (see Chapter 4) suggest that nonperipheral vowels are more likely to be perceptually confused than peripheral vowels. Thus, according to McGahay (2024), Turkish vowel system should go through 'iterative confusion minimization' whereby the vowels that are highly perceptually confusable change their phonetic targets to become less perceptually confusable until optimal dispersion is achieved in the vowel system. Although I do not have time series data to examine whether the nonperipheral vowels in Turkish have become less perceptually confusable in time, I argue that that vowel cooccurrence constraints are stronger for nonperipheral vowels suggests that Turkish vowel harmony might be a way to optimize the Turkish vowel system without increasing dispersion in the vowel space through sound change. This claim follows from my observation that Turkish vowels that are perceptually highly confusable have lower likelihood of cooccurrence in the lexicon. I argue that the lexicon is organized in a way that contextually reduces the probability with which nonperipheral vowels will be perceptually confused. Due to this lower perceptual confusability of vowels in the lexicon, the push for a sound change to attain optimal dispersion might not arise. Future studies might revise the sound change simulations presented in McGahay to integrate the likelihood of vowel cooccurrence in the lexicon as a predictor to examine whether a suboptimal vowel system might survive with the help of vowel harmony, as I propose might be the case for Turkish.

The main limitation in this study was that the dataset of perceptual confusions in the Turkish vowel identification dataset was small, considering I was interested in perceptual confusions in vowel identification as a measure of perceived vowel similarity. In analyzing the Turkish vowel identification responses in Chapter 4, I ran into model convergence issues which led me to exclude random effects structures and convert individual responses to count data. As a result, my analyses do not fully capture the variation in the data by stimulus items or participants. This limitation could be addressed with a larger dataset. In this experiment, there was one token per each of the 8 Turkish vowels produced by 20 speakers, totaling 160 Turkish vowel identification stimuli. The full dataset consisted of 6400 datapoints, and the dataset of perceptual confusions had less than 700 datapoints. In contrast, the English vowel identification experiment by Hillenbrand et al. (1995) had a total of 1668 stimuli consisting of 12 English vowels produced by 139 speakers. Their English vowel identification dataset was 5 times larger, leading to a higher number of perceptual confusions that can be modeled. In future studies, vowel productions from a larger number of native Turkish speakers could be elicited to be as stimuli in Turkish vowel identification experiments to increase the number of datapoints and yield a higher number of perceptual confusions that can be modeled with more precision.

Overall, the studies presented in this dissertation provide partial support for the exemplar models hypothesis that higher likelihood of vowel cooccurrence leads to higher perceived vowel similarity. I find that this hypothesis is qualified by perceptual confusability of vowels and in particular weak phonetic contrasts. The post-hoc analyses contribute to previous discussions on vowel peripherality as a factor influencing vowel cooccurrence restrictions in Turkish beyond agreement in vowel features. As suggested by Kabak and Weber (2013), Turkish vowel harmony as it applies in the lexicon has not received extensive attention in the literature. In line with their observations, I argue that a lexically grounded approach deepens our understanding of Turkish vowel harmony and reveals vowel cooccurrence patterns and relationships that otherwise might go unnoticed. Lastly, to my knowledge, this dissertation is the first to present evidence from

Turkish vowel perception in support of Suomi's (1983) claim that Turkish vowel harmony is a perceptually motivated phenomenon. Future work might put this claim to test more directly.

References

Aksu, B. (2023). Türkçe Sesbilim Araştırmaları için Kapsamlı Bir Metin Arayışı: "Rüzgar ile Güneş" ve "Yalancı Çoban" Metinlerinin Karşılaştırılması. *Akademik Dil ve Edebiyat Dergisi*, *7*(1), 135–154. https://doi.org/10.34083/akaded.1256702

Babel, M., & Johnson, K. (2010). Accessing psycho-acoustic perception and language-specific perception with speech sounds. *Laboratory Phonology*, *1*(1), 179–205. https://doi.org/10.1515/labphon.2010.009

Bakay, Ö., Ergelen, Ö., Sarmış, E., Yıldırım, S., Kocabalcıoğlu, A., Arıcan, B. N., … Yıldız, O. T. (2021). Turkish WordNet KeNet. In P. Vossen & C. Fellbaum (Eds.), *Proceedings of the 11th Global Wordnet Conference* (pp. 166–174). Global Wordnet Association. https://doi.org/https://doi.org/978-946402731-0

Baker, A. C. (2009). Two Bayesian Approaches to Finding Vowel Harmony.

Barnes, J. (2002). Domain-initial strengthening and the phonetics and phonology of positional neutralization. *Proceedings of North East Linguistic Society 32*, *1*, 1–20.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Beddor, P. S., Harnsberger James D., J. D., & Lindemann, S. (2002). Language-specific patterns of vowel-to-vowel coarticulation: Acoustic structures and their perceptual correlates. *Journal of Phonetics*, *30*(4), 591–627. https://doi.org/10.1006/jpho.2002.0177

Beddor, P. S., & Yavuz, H. K. (1995). The relation between vowel-to-vowel coarticulation and vowel harmony in Turkish. In *Proceedings of the 13th International Congress of Phonetic Sciences* (pp. 44–51).

Bellik, J. (2018). An acoustic study of vowel intrusion in Turkish onset clusters. *Laboratory*

Phonology: Journal of the Association for Laboratory Phonology, *9*(1), 16. https://doi.org/10.5334/labphon.112

Boomershine, A., Hall, K. C., Hume, E., & Johnson, K. (2008). The impact of allophony versus contrast on speech perception. In *Contrast in Phonology: Theory, Perception, Acquisition* (pp. 145–172). https://doi.org/10.1515/9783110208603.2.145

Börtlü, G. (2020). *The vowel triangle of Turkish and phonological processes of laxing and fronting in Turkish*. Hacettepe University.

Boyce, S. E. (1990). Coarticulatory organization for lip rounding in Turkish and English. *The Journal of the Acoustical Society of America*, *88*(6), 2584–2595. https://doi.org/10.1121/1.400349

Canalis, S., & Dikmen, F. (2020). Turkish palatalized consonants. *Proceedings of the Workshop on Turkic and Languages in Contact with Turkic*, *5*(1), 41–55. https://doi.org/10.3765/ptu.v5i1.4781

Cangökçe Yaşar, Ö., Kızıltan, N., & Baydal, D. (2019). "Jale'nin Dünyası" Metninin Ses Zenginliği Bağlamında İncelenmesi. In K. İşeri (Ed.), *Dilbilimde Güncel Tartışmalar*. Dilbilim Derneği Yayınları.

Caplan, S., & Kodner, J. (2018). The Acquisition of Vowel Harmony from Simple Local Statistics. In T. T. Rogers, M. Rau, X. Zhu, & C. W. Kalish (Eds.), *Proceedings of the 40th Annual Meeting of the Cognitive Science Society, CogSci 2018* (pp. 1440–1445). Cognitive Science Society.

Carignan, C. (2018). Using ultrasound and nasalance to separate oral and nasal contributions to formant frequencies of nasalized vowels. *The Journal of the Acoustical Society of America*, *143*(5), 2588–2601. https://doi.org/10.1121/1.5034760

Chen, M. Y. (1995). Acoustic parameters of nasalized vowels in hearing-impaired and normal-hearing speakers. *The Journal of the Acoustical Society of America*, *98*(5), 2443–2453. https://doi.org/10.1121/1.414399

Cho, T. (2005). Prosodic strengthening and featural enhancement: Evidence from acoustic and articulatory realizations of /ɑ,i/ in English. *The Journal of the Acoustical Society of America*, *117*(6), 3867–3878. https://doi.org/10.1121/1.1861893

Clements, G. N., & Sezer, E. (1982). Vowel and Consonant Disharmony in Turkish. In *Linguistic Models 3* (pp. 213–256). De Gruyter. https://doi.org/10.1515/9783112423325-007

Cole, J. (2009). Emergent feature structures: harmony systems in exemplar models of phonology. *Language Sciences*, *31*(2–3), 144–160. https://doi.org/10.1016/j.langsci.2008.12.004

Coşkun, V. (2000). Türkiye ve Özbek Türkçelerinde "ğ." *Türk Dünyası Dil ve Edebiyat Dergisi*, *9*(Bahar), 251–260.

Dadan, M. C., Erken, U., & Canalis, S. (2024). /e/ lowering in Turkish. In *Proceedings of the Workshop on Turkic and Languages in Contact with Turkic 9*. https://doi.org/https://doi.org/10.3765/gk0kay89

Davidson, L., & Shaw, J. A. (2012). Sources of illusion in consonant cluster perception. *Journal of Phonetics*, *40*(2), 234–248. https://doi.org/10.1016/j.wocn.2011.11.005

Ergenç, İ., & Uzun, İ. P. (2020). *Türkçenin ses dizgesi*. Seçkin Yayıncılık.

Erguvanlı-Taylan, E. (2015). *The Phonology and Morphology of Turkish*. Boğaziçi University Press.

Ettlinger, M., & Johnson, K. (2010). Vowel Discrimination by English, French and Turkish Speakers: Evidence for an Exemplar-Based Approach to Speech Perception. *Phonetica*, *66*(4), 222–242. https://doi.org/10.1159/000298584

Foster, S., & Cole, J. (2024). Trading Relations in Segmental Cues to Prosodic Prominence. In *Speech Prosody 2024* (pp. 577–581). ISCA: ISCA.

https://doi.org/10.21437/SpeechProsody.2024-117

Fox, R. A. (1984). Effect of Lexical Status on Phonetic Categorization, *10*(4), 526–540.

Fry, D. B., Abramson, A. S., Eimas, P. D., & Liberman, A. M. (1962). The Identification and Discrimination of Synthetic Vowels. *Language and Speech*, *5*(4), 171–189. https://doi.org/10.1177/002383096200500401

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, *6*(1), 110–125. https://doi.org/10.1037/0096-1523.6.1.110

Gerrits, E., & Schouten, M. E. H. (2004). Categorical perception depends on the discrimination task. *Perception & Psychophysics*, *66*(3), 363–376. https://doi.org/10.3758/BF03194885

Getz, L. M., & Toscano, J. C. (2021). The time-course of speech perception revealed by temporally-sensitive neural measures. *Wiley Interdisciplinary Reviews: Cognitive Science*, *12*(2), 1–18. https://doi.org/10.1002/wcs.1541

Göksel, A., & Kerslake, C. (2005). *Turkish: A Comprehensive Greammar*. Routledge.

Gopal, D., & Nichols, S. (2022). Apparent-time change in Turkish front-mid-vowel allophony.

Güngör, T. (2003). Lexical and morphological statistics for Turkish. In C. Ardıl (Ed.), *Proceedings of the International Twelfth Turkish Symposium on Artificial Intelligence and Neural Networks (TAINN-2003)* (pp. 409–412).

Hall, K. C., & Hume, E. (2015). Modeling Perceived Similarity: The Influence of Phonetics, Phonology and Frequency on the Perception of French Vowels.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*, *97*(5), 3099–3111. https://doi.org/10.1121/1.411872

Inkelas, S., Barnes, J., Good, J., Kavitskaya, D., Orgun, O., Sprouse, R., & Yu, A. (2001). Stress and vowel-to-vowel coarticulation in Turkish. *Annual Meeting of the Linguistic Society of America*.

Iverson, P., & Kuhl, P. K. (2000). Perceptual magnet and phoneme boundary effects in speech perception: Do they arise from a common mechanism? *Perception and Psychophysics*, *62*(4), 874–886. https://doi.org/10.3758/BF03206929

Jackman, S. (2024). pscl: Classes and Methods for R Developed in the Political Science Computational Laboratory. Retrieved from https://github.com/atahk/pscl/

Johnson, K. (1997). Speech perception without speaker normalization: an exemplar model. In K. Johnson & J. W. Mullennix (Eds.), *Talker Variability in Speech Processing* (pp. 145–165).

Johnson, K. (2015). Vowel Perception Asymmetry in Auditory and Phonemic Listening. *UC Berkeley Phonology Lab Annual Reports*, *11*. https://doi.org/10.5070/p721t337gh

Johnson, K., & Babel, M. (2010). On the perceptual basis of distinctive features : Evidence from the perception of fricatives by Dutch and English speakers. *Journal of Phonetics*, *38*(1), 127–136. https://doi.org/10.1016/j.wocn.2009.11.001

Kabak, B. (2007). Hiatus resolution in Turkish: An underspecification account. *Lingua*, *117*(8), 1378–1411. https://doi.org/10.1016/j.lingua.2006.04.010

Kabak, B. (2011). Turkish Vowel Harmony. *The Blackwell Companion to Phonology*, 1–24. https://doi.org/10.1002/9781444335262.wbctp0118

Kabak, B., Kasselkus, E., Maniwa, K., & Weber, S. (2008). Vowel harmony has direction and context: a corpus study. In *16th Manchester Phonology Meeting* (pp. 22–24).

Kabak, B., & Weber, S. (2013). Markedness, context, and directionality in vowel harmony: A corpus study on vowel co-occurrence patterns.

Kaun, A. R. (1995). *The typology of rounding harmony: An Optimality Theoretic approach*. University of California, Los Angeles.

https://doi.org/https://doi.org/doi:10.7282/T3R49PM2

Kaun, A. R. (2004). The typology of rounding harmony. *Phonetically Based Phonology*, (Svantesson 1985), 87–116. https://doi.org/10.1017/CBO9780511486401.004

Kılıç, M. A., & Erdem, M. (2013). Türkiye Türkçesindeki "yumuşak g" ünsüzünün fonetik analizi. In *VI. Uluslararası Türk Dili Kurultayı Bildirileri* (pp. 2809–2826).

Kılıç, M. A., & Öğüt, F. (2004). A high unrounded vowel in Turkish: Is it a central or back vowel? *Speech Communication*, *43*(1–2), 143–154. https://doi.org/10.1016/j.specom.2004.03.001

Koo, H., & Cole, J. (2007). Gradient perceptual facilitation from phonotactic knowledge. *Proceedings of the 43rd Annual Meeting of the {Chicago Linguistic Society}*, (January 2007), 167–181.

Kopkallı-Yavuz, H. (2010). The Sound Inventory of Turkish: Consonants and Vowels. In *Communication Disorders in Turkish* (pp. 27–47). Channel View Publications. https://doi.org/10.2307/jj.29308444.9

Kornfilt, J. (1997). *Turkish*. Routledge.

Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, *50*(2), 93–107. Retrieved from https://link.springer.com/content/pdf/10.3758/BF03212211.pdf

Kunduracı, A. (2013). Sesbilim-Biçimbilim Arakesiti. In İ. P. Uzun (Ed.), *Kuramsal ve uygulamalı sesbilim*. Seçkin Yayıncılık.

Labov, W. (1994). *Principles of Linguistic Change: Internal Factors*. Blackwell.

Labov, W., Yaeger, M., & Steiner, R. (1972). *A Quantitative Study of Sound Change in Progress*.

Lees, R. B. (1961). The so-called soft-g. In *The phonology of Modern Standard Turkish* (pp. 7–9). Indiana University Publications.

Lees, R. B. (1966). On the Interpretation of a Turkish Vowel Alternation. *Anthropological Linguistics*, *8*(9), 32–39.

Lenth, R. V., & Piaskowski, J. (2025). emmeans: Estimated Marginal Means, aka Least-Squares Means.

Levi, S. V. (2001). Glides, laterals, and Turkish Vowel Harmony. In *CLS 37* (pp. 379–393).

Levi, S. V. (2004). *The representation of underlying glides: a cross-linguistic study*.

Liberman, A. M., Harris, K. S., Howard, S., Griffith, B. C., & Hill, M. (1957). The discrimination of speech sounds within and across phoneme boundaries, *54*(5), 358–368.

Manuel, S. Y. (1990). The role of contrast in limiting vowel-to-vowel coarticulation in different languages. *The Journal of the Acoustical Society of America*, *88*(3), 1286–1298. https://doi.org/10.1121/1.399705

Masapollo, M., Polka, L., Molnar, M., & Ménard, L. (2017). Directional asymmetries reveal a universal bias in adult vowel perception. *The Journal of the Acoustical Society of America*, *141*(4), 2857–2869. https://doi.org/10.1121/1.4981006

Mayer, T., Rohrdantz, C., Butt, M., Plank, F., & Keim, D. A. (2010). Visualizing Vowel Harmony. *Linguistic Issues in Language Technology*, *4*(December). https://doi.org/10.33011/lilt.v4i.1223

McGahay, J. F. (2024). *Vowel System Typology and Confusion Minimization*. Retrieved from https://escholarship.org/uc/item/48s9011k

McMurray, B. (2022). The myth of categorical perception. *The Journal of the Acoustical Society of America*, *152*(6), 3819–3842. https://doi.org/10.1121/10.0016614

Oded, I., Idsardi, W. J., & Rhone, A. (2008). Processing of vowel (dis) harmony in Turkish. *Studia Uralo-Altaica*, *49*, 361–370. Retrieved from https://ojs.bibl.u-szeged.hu/index.php/stualtaica/article/view/13687

Özsoy, A. S. (2004). *Türkçenin Yapısı - I: Sesbilim*. Boğaziçi Üniversitesi Yayınevi.

Peterson, G. E., & Barney, H. L. (1952). Control Methods Used in a Study of the Vowels. *Journal of the Acoustical Society of America*, *24*(2), 175–184. https://doi.org/10.1121/1.1906875

Pierrehumbert, J. B. (2001). Exemplar dynamics. In J. L. Bybee & P. J. Hopper (Eds.), *Frequency and the Emergence of Linguistic Structure* (pp. 137–158). John Benjamins Publishing Company. https://doi.org/10.1075/tsl.45.08pie

Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels *. *Perception and Psychophysics*, *13*(2), 253–260.

Pisoni, D. B., & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics*, *15*(2), 285–290. https://doi.org/10.3758/BF03213946

Polka, L., & Bohn, O.-S. (2003). Asymmetries in vowel perception. *Speech Communication*, *41*(1), 221–231. https://doi.org/10.1016/S0167-6393(02)00105-X

Polka, L., & Bohn, O.-S. (2011). Natural Referent Vowel (NRV) framework: An emerging view of early phonetic development. *Journal of Phonetics*, *39*(4), 467–478. https://doi.org/10.1016/j.wocn.2010.08.007

Radisic, M. (2014). An Ultrasound and Acoustic Study of Turkish Rounded / Unrounded Vowel Pairs.

Rizzi, R., & Bidelman, G. M. (2024). Functional benefits of continuous vs. categorical listening strategies on the neural encoding and perception of noise-degraded speech. *Brain Research*, *1844*(May), 149166. https://doi.org/10.1016/j.brainres.2024.149166

Şayli, Ö. (2002). *Duration analysis and modelling for Turkish text-to-speech synthesis*. Boğaziçi University.

Schouten, B., Gerrits, E., & Van Hessen, A. (2003). The end of categorical perception as we know it. *Speech Communication*, *41*(1), 71–80. https://doi.org/10.1016/S0167-6393(02)00094-8

Schwartz, J. L., Abry, C., Boë, L. J., Ménard, L., & Vallée, N. (2005). Asymmetries in vowel perception, in the context of the Dispersion- Focalisation Theory. *Speech Communication*, *45*(4), 425–434. https://doi.org/10.1016/j.specom.2004.12.001

Suomi, K. (1983). Palatal Vowel Harmony: A Perceptually Motivated Phenomenon? *Nordic Journal of Linguistics*, *6*(1), 1–35. https://doi.org/10.1017/S0332586500000949

Taşdemir, B., Cahill, G., & Elfner, E. (2025). Revisiting the “Soft g” <ğ> in Standard Turkish: Phonetic Evidence and Historical Orthographic Challenges. In *Proceedings of the Workshop on Turkic and Languages in Contact with Turkic Tu+ 10*.

Terbeek, D. (1977). A cross-language multidimensional scaling study of vowel perception.

Toscano, J. C., Anderson, N. D., Fabiani, M., Gratton, G., & Garnsey, S. M. (2018). The time-course of cortical responses to speech revealed by fast optical imaging. *Brain and Language*, *184*(June), 32–42. https://doi.org/10.1016/j.bandl.2018.06.006

Türk, O., Şayli, Ö., Özsoy, A. S., & Arslan, L. M. (2004). Türkçe’de Ünlülerin Formant Analizi. In *18. Dilbilim Kurultayı*.

Ünal-Logacev, Ö., Zygis, M., & Fuchs, S. (2019). Phonetics and phonology of soft “g” in Turkish. *Journal of the International Phonetic Association*, *49*(2), 183–206. https://doi.org/10.1017/S0025100317000317

Yavaş, M. (1980). Some pilot experiments on Turkish vowel harmony. *Paper in Linguistics*, *13*(3), 543–562. https://doi.org/10.1080/08351818009370510

Yılmaz, C. (2022). *Türkçe ünlülerin algılanmasında süre, temel frekans ve spektral tilt etkisi*. Anadolu Üniversitesi.

Yılmaz Davutoğlu, A. (2010). *Standart Türkçedeki ünlülerin akustik analizi ve fonetik altyapı*. İstanbul Üniversitesi.

Zeileis, A., Kleiber, C., & Jackman, S. (2008). Regression models for count data in R. *Journal of Statistical Software*, *27*(8), 1–25. https://doi.org/https://doi.org/10.18637/jss.v027.i08

Zimmer, K. (1969). Psychological Correlates of Some Turkish Morpheme Structure Conditions. *Language*. https://doi.org/10.2307/411662

Zimmer, K., & Küntay, A. (2003). Turkish internal vowel harmony revisited. In A. S. Özsoy, D. Akar, M. Nakipoğlu-Demiralp, E. Erguvanlı-Taylan, & A. Aksu-Koç (Eds.), *Studies in Turkish Linguistics* (pp. 27–33). Boğaziçi University Press.

Zimmer, K., & Orgun, O. (1992). Turkish. In *Handbook of the International Phonetic Association* (pp. 2–4).