

UC Davis

UC Davis Electronic Theses and Dissertations

Title

Vowel Length in German: Use of Quality and Quantity in the Perception of Long and Short Vowels in German

Permalink

<https://escholarship.org/uc/item/04q903t4>

Author

Predeck, Kristin

Publication Date

2022

Peer reviewed|Thesis/dissertation

Vowel Length in German:
Use of Quality and Quantity in the Perception of
Long and Short Vowels in German

By
KRISTIN PREDECK
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

German

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Carlee Arnett, Co-Chair

Santiago Barreda, Co-Chair

Georgia Zellou

Committee in Charge

2022

Table of contents

Chapter 1	...1
1.1 Introduction	...1
1.2 Literature Review	... 5
1.2.1 A short history of German	...6
1.2.2 Quantity in German - tense/lax or long/short?	...16
1.2.2.2 Using acoustic evidence to investigate quantity and quality	...20
1.2.3 The quantity-quality debate	...26
1.3 Theories of Speech Perception	...31
1.3.1 Gestural Theories	...33
1.3.2 Segmental Theories	...36
1.3.3 Auditory/Acoustic theories	...37
1.3.4 Exemplar Theories	...38
1.4 Normalization and the invariance problem	...39
1.4.1 Extrinsic Theories	...40
1.4.2 Vowel normalization as perceptual constancy	...43
1.5 Approach used for this dissertation	...45
Chapter 2	...46
2.1 Experiment One	...47
2.1.1 Background	...49
2.1.2 SLA theories and model	...51
2.1.3 CA English vowels	...56
2.2 Materials and Methods	...58
2.2.1 Participants	...58
2.2.1.1 German	...58
2.2.1.1 English	...59
2.2.2 Production stimuli	...59
2.2.2.1 German	...59
2.2.2.2 English	...60
2.2.3 Procedure	...60
2.2.3.1 Production	...60
2.2.3.1.1 German	...60
2.2.3.1.2 English	...61
2.2.3.2 Perception	...6

2.2.4 Acoustic measures	...65
2.2.6 Results	...67
2.2.6.1 Production	...67
2.2.6.1.1 German	...67
2.2.6.1.2 English	...70
2.2.6.2 Perception	...72
2.2.6.2.1 Human Listener Perception	...72
2.2.6.2.2 Vowel Specific Models	...76
2.2.7 Discussion	...81
Chapter 3	...87
3.1 Experiments two and three	...87
3.2 Experiment two - Duration Continua	...87
3.2.1 Introduction	...87
3.2.2 Methods and Materials	...88
3.2.2.1 Listeners	...88
3.2.2.2 Stimuli	...89
3.2.2.3 Procedure	...89
3.2.3 Analysis	...91
3.2.4 Results	...91
3.2.5 Interim Discussion	...103
3.3 Experiment three - Spectral Continua	...104
3.3.1 Methods and Materials	...105
3.3.1.1 Listeners	...105
3.3.1.2 Stimuli	...105
3.3.1.3 Procedure	...106
3.3.2 Analysis	...107
3.3.3 Results	...108
3.3.4 Interim Discussion	...114
3.4 General Discussion	...116
Chapter 4	...119
4.1 Introduction	...119
4.2 Experiment four	...122
4.2.1 Methods and Materials	...123
4.2.1.1 Listeners	...123

4.2.1.2 Stimuli	...124
4.2.1.2.1 Production	...124
4.2.1.2.2 Resynthesized Productions	...125
4.2.2 Procedure	...129
4.2.3. Analysis	...131
4.2.4 Results	...132
4.3 Discussion	...135
Chapter 5	...138
5.1 General Conclusion	...138
References	...144
Appendix	...166

COVID-19 Statement

Due to the current COVID-19 situation, all experiments were presented in Qualtrics and completed from the participants' homes. Due to this, the quality of recordings was not the same as it would have been in a sound attenuated booth in a lab. Additionally participants had to rely on their own understanding of the instructions and were not able to ask questions.

Tools used

All data analysis, cleaning, and manipulation were done on the author's personal Macbook Pro. Acoustic features were extracted in Praat. Stimulus resynthesis was done in Python. Acoustic and statistical analyses were done in Python using the sklearn package and plots were made in R using the PhonTools (Barreda 2015) and PhonR (McCloy 2016) packages.

Acknowledgements

My sincerest gratitude goes to my advisors, Carlee Arnett and Santiago Barreda for their continued support, guidance, and encouragement. This dissertation would not have been possible without them. I am also sincerely grateful to my third committee member, Georgia Zellou for always being supportive and getting me her comments in record time. My dissertation has greatly benefitted from my committee members' thoughtful suggestions and comments.

I also want to thank my fellow graduate student Aleese Block, who has taught me to be confident, has provided me much needed emotional support, and has lent me her ear more than once during the writing of this dissertation. I also thank my fiancé, Robert, for being patient and listening to my worries and anxieties with unlimited patience.

I am also incredibly grateful to the individuals who participated in this study giving me their own time without expecting anything in return. Specifically, I am thankful to my dad, Frank, for piloting all my experiments multiple times without complaint. I am also grateful to my grandfather, Manfred, who sadly passed away before he could see me graduate. The memories I have of him and me conducting little experiments in his lab and sharing his own struggles in his career, but never giving up, have given me the inspiration and strength to finish this dissertation. Danke, Opi!

And lastly I owe thanks to Wesley Brooks for taking the time to walk through the pros and cons of different statistical methods with me and patiently explaining Bayesian principles and model outputs to me.

Abstract

The acoustic cues used to distinguish different vowel categories from each other differ from language to language. While native speakers of a language are well attuned to the important cues needed to identify different sounds with a high degree of accuracy, language specific cues and cue weightings can cause potential problems for second language learners because of the learned L1 contrasts. German is a language that shows quantity in its vowel system with every long vowel having a corresponding short vowel. It is unclear, however, whether spectral information or duration is used as the primary cue by German listeners to disambiguate whether a vowel is short or long. Additionally, the quantity-quality debate suggests that only one of the two features is distinctive while the other one is redundant, ignoring the potential use of secondary cues. This dissertation seeks to provide a clear answer to whether quality or quantity is used as a primary cue in vowel perception in German as well as the potential use of secondary in the disambiguation of long short vowel pairs.

Results show that American English listeners perceived all German vowel pairs as different native categories, with the exception of /ɛ:/-/ɛ/ and /u:/-/ʊ/, largely relying on spectral features when identifying non-native vowel sounds. Therefore German vowels show substantial spectral differences in production that are salient enough for non-native listeners to exploit. In native speech perception, results show that while duration is used as a primary cue, spectral information was used as a secondary cue to disambiguate whether a vowel was long or short. While listeners identified tokens as long less often as tokens approached short durations, they still identified those tokens containing originally long spectral information as long more often than those containing short spectral information. The same patterns were found in the second experiment, with duration being used as the primary cue in disambiguating whether a vowel was long or short. Tokens containing originally short durations were selected as long less than 50% of the time, regardless of spectral manipulation step. Additionally, German listeners used spectral movement at least partially. Results from experiment four show that while identification accuracy stayed high overall, results indicate that German listeners rely on dynamic information, with silent-onset-offset tokens having lower identification accuracies than the silent-middle tokens. This is further evidence for spectral information being used as a secondary cue in the disambiguation of German long and short vowel pairs.

Chapter 1

1.1 Introduction

An acoustic cue can be broadly described as “information in the acoustic signal that allows the listener to apprehend the existence of a phonological contrast.” (Wright 2004:36). Vowels differ on multiple phonetic dimensions and languages differ in which cues are used in speech perception and in the relevance of the individual cues. Vowel quality has been strongly linked to the first two formant frequencies and it is used to describe a simple target model of perception in which the first two formants (F1 and F2) constitute the primary source of acoustic information needed in successful vowel discrimination (c.f., Peterson 1961, Nearey 1978, Peterson and Barney 1952). Some languages additionally rely on duration, such as German (c.f. Tomaschek 2011, 2013, 2014, Bennett 1968, Bohn and Polka 2001) and Swedish (c.f. Behne et al. 1996, Behne et al. 1997, Elert 1964, Hadding-Koch and Abramson 1964), to disambiguate vowels. For example, the German words *Kamm* /kam/ (English *comb*) and *kahm* /ka:m/ (English *came*) use vowel length to distinguish two different lexical meanings. There are few studies looking into the combination of how duration and spectral information are used in the perception of German long-short vowel pairs. Instead, studies have focused on either the importance of duration (c.f. von Essen 1979, Heike 1969, 1970, 1972, Lindner 1976, Weiss 1976, Sendlmeier 1981) or spectral information (c.f. Bennett 1968, Ungeheuer 1969, Strange and Bohn 1998) as a primary cue and do not discuss how spectral information and duration are used in combination.

Additionally, some studies have shown F3 and f0 to be important in vowel classification, at least indirectly (c.f. Nearey 1989, Slawson 1968, Hillenbrand and Gayvert 1993, Glidden and Assmann 2004, Assmann and Nearey 2007), so relying solely on either F1 and F2 from the midpoints or duration values of the vowels seems to be insufficient when describing the reality of vowel production and perception. The work of Peterson and Barney (1952) has shown substantial variance and overlap of vowel categories in the target frequencies of F1 and F2 in the speech of men, women, children, and even speakers of the same age and gender.

In this dissertation, a total of four experiments looked into different aspects of German vowel production and perception in order to investigate the role of duration and the first three formant frequencies in the contrasts of German long-short vowel pairs. Specifically, the focus was on vowels as acoustic rather than articulatory events. How do listeners use the acoustic cues present in the signal to recover the speaker's intended phoneme? A series of four experiments was conducted to answer this question.

The first experiment set out to investigate the production of German vowels and English vowels, and the perception of German vowels by English listeners. The identification of non-native sounds depends on the phonetic fit and phonological closeness to the L1 phonemes (Flege 1995). The experiment establishes the acoustic qualities present in the production of German vowels and the saliency of these cues in perception by naive listeners. This is an important step to establish the saliency of acoustic cues present in German long short vowels. English is a closely related language to German but has lost the quantity contrasts present in German over time.

Additionally, the vowel system is also smaller and shows less overlap spectrally, which is why American English listeners are likely to pick up on spectral differences between German long short vowel pairs. If these vowels differ mainly based on duration, naive listeners might collapse them into the same native English category. However, if spectral differences are salient, listeners are more likely to exploit these differences and perceive the long and short vowels as distinct and therefore map them to different American English vowels. The data from experiment one, therefore, provides insight into both the contrastive phonetic properties of both American English and German and how much speech perception is constrained by the acoustic patterns of the L1.

One way to examine perceptual cues is to manipulate one acoustic cue at a time to see what listeners are most sensitive to (Cooper, Liberman, Borst 1951). This method is used for experiments two and three to investigate whether spectral cues or duration is the primary cue used in the disambiguation of long-short vowel pairs in German. The second experiment investigates the role of temporal information in the perception of German long-short vowels more closely. In experiment two a five-step continuum of durations was created, going from formant durations of originally long vowels to durations of originally short vowels. Two sets of these duration continua were synthesized, with one set keeping the formant frequencies of the originally long vowel, and one set keeping the formant frequencies of the originally short vowel intact. If German listeners rely mainly on duration, they should perceive tokens as long as they approach the long duration, regardless of original formant information. The third experiment investigates the role of spectral information in speech perception more closely. German listeners were presented with a five-step continuum of manipulated

F1/F2/F3 values going from formant frequencies of originally long vowels to formant frequencies of originally short vowels. Two sets of these formant continua were synthesized with one set keeping the duration of the original long vowel and one set keeping the duration of the originally short vowel intact. If German listeners rely mainly on spectral information, they should perceive tokens as long as they approach the long spectrum, regardless of original duration information.

A fourth experiment was conducted to investigate the importance of formant movement. Listeners were presented with silent center/silent onset-offset resynthesized vowels to test the importance of vowel inherent spectral change (VISC). Traditionally German is thought to rely less on dynamic and more on static cues as it is said to not diphthongize monophthongs in comparison to American English (Strange and Bohn 1998, Strange et al. 2004). However, German has a large vowel system and shows spectral overlap between vowel categories. Using information from spectral trajectories might be a way of further disambiguating vowel categories. If German listeners rely mainly on static spectral information, the silent-center tokens should have lower identification accuracies than the silent-onset-offset tokens. However, if German listeners rely on dynamic information, the identification accuracies should be lower in the silent-onset-offset tokens.

Furthering the understanding of cue usage in speech perception not only aids the application of appropriate teaching materials in second language acquisition classes but also allows for the development of algorithms and methods for automatic speech recognition modeling more closely what human listeners are doing in a specific language. Modifying existing algorithms to more accurately classify unfamiliar speech

sounds can be aided by a better understanding of both native and cross-language human perception. Additionally, incorporating information from fine-grained acoustic studies can lead not only to greater accessibility of speech technologies for dialects of one language and accented speech, such as in second language speakers, it could also lead to the development of other more inclusive technologies, for example, the availability of speech recognition technology for speech impaired speakers and the development of hearing aids that pick up on the right cues.

1.2 Literature review

In the following sections, a short overview of the literature on German vowel characteristics and American English vowel characteristics, as well as speech perception will be given. This section will be the theoretical basis on which the experiments are built. The information in this chapter will be important for experiments one through four, which will investigate the use of durational and spectral information in German vowel production and perception using manipulated synthetic vowels.

After establishing the acoustic principles of vowel production, a short overview of speech perception and vowel normalization will be given. Understanding how non-native speakers use acoustic cues to categorize unfamiliar sounds can give insight into the distribution of acoustic properties in the native language and how categories are established based on the properties of the signal.

1.2.1 A short history of German

“Germania tot habet dialectos, ut in triginta miliaribus homines se mutuo non intelligant. Austri et Bavari nullas servant diphthongos, dicunt enim e ur, fe ur, bro edt pro feuer, euer, brodt. Ita Francones unisona et crassa voce loquuntur, quod Saxones praeq̄ipue Antverpiensium linguam non intelligunt [...] die Oberlendische sprache ist nicht die rechte Teutzsche sprache, habet enim maximos hiatus et sonitus, sed Saxonica lingua est facillima, fere pressis labiis pronunciatur.”

Germany has so many dialects that people living 30 miles from each other can not understand each other. The Austrians and Bavarians do not have diphthongs, because they say e-ur, fe-ur, bro-edt for feuer, euer, brodt. The Franks speak so monotonous and thick, that the Saxons do not understand them, especially in Antwerp [...] Oberlendisch language is not the right German language, because it has very open and strong sounds, but the Saxon language is very light, it is pronounced with almost closed lips.

(from Luther’s “Tischreden”, Lauterbachs Sammlung B, Nr. 6146, in: Stedje 2007:146.)

While English and German both descended from West-Germanic, they went through a series of different processes of sound change, which is also reflected in the current vowel systems of both; while German has 23 monophthongs and 8 diphthongs, English has only 12 monophthongs and 8 diphthongs (Calvo Fernández 2018). The smaller vowel system of English shows less overlap in the F1/F2 plane than German does, which could explain why German went through sound change processes that preserved duration as a strong perceptual cue (D’Alquen 1979). Comparing the vowel systems of American English and German allows insight into the acoustic cues present in vowel production and the active phonological contrasts in perception in both languages. With English having a smaller vowel inventory and less overlap spectrally, the attention to spectral form versus duration might be inverse, with English speakers relying more on

spectral form and German speakers relying more on duration in the perception of vowel sounds.

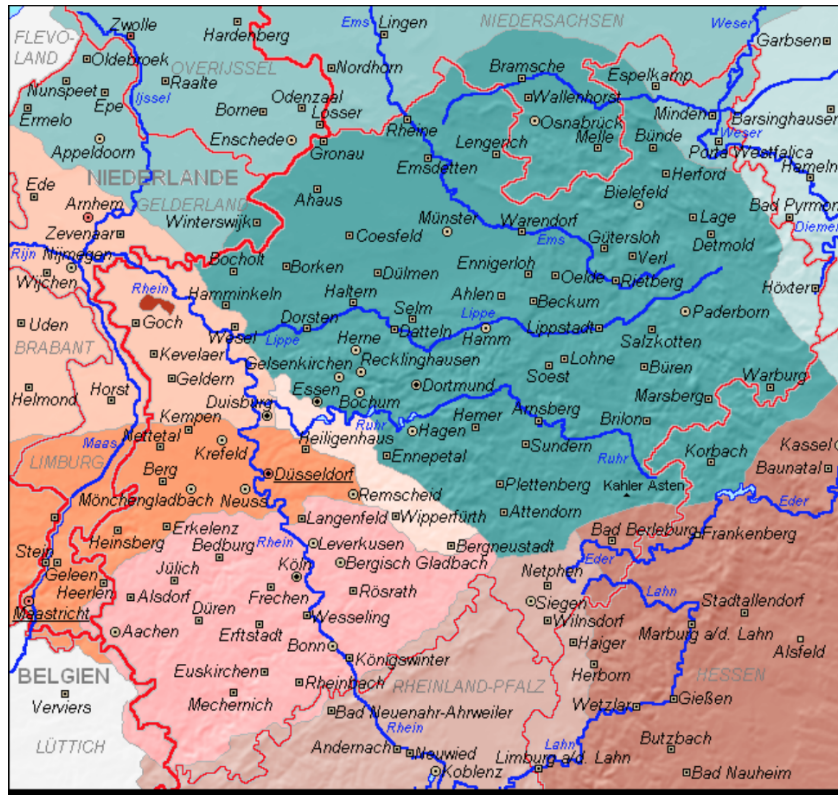
In this dissertation, the focus of study will be on the vowels of one German dialect, Westphalian. The reason for focusing on a specific dialect is that Modern High German or Modern Standard German (MSG) as such is a construct (c.f. Lenz 2001, Milroy 2007), as German has many dialects and MSG is not actually spoken in everyday practice (c.f. Barbour and Stevenson 1990). While in written language, uniformity can be closely approximated, spoken language is never completely invariant and absolute unity is never achieved in practice (Milroy 2007:134). The standard of a language is not a specific or clear-cut language but rather “an idea in the mind rather than a reality – a set of abstract norms to which actual usage may conform to a great or lesser extent” (Milroy and Milroy 1991:22-23) and the more prestigious dialects of a language often rise to be the standard. Unlike French or English, MSG cannot be traced back to one specific dialect, mainly due to the political fragmentation of Germany between the 16th century and the second half of the 19th century (Barbour & Stevenson 1990, Clyne 1995, Zsiga 2013), which prevented any one region from becoming prestigious enough to be considered to become the standard (c.f. Barbour & Stevenson 1990, Clyne 1995, Zsiga 2013). Because of Germany’s history of scattered regionalism and the lack of a language authority or institution deciding on a norm, the disunity of spoken German is still present today and a phonetic norm was only established in the 19th century (*Hochlautung*) (Keller 1978). The *Hochlautung* is not spoken in everyday life in Germany as regional dialects have persisted and while a single speaker’s speech may conform to a certain extent to the idealized standard of a language, variation in the

acoustic realizations will still be present (c.f. Lippi-Green 1997:25, Stevenson 2002, Kennetz 2010).

There are multiple problems caused by the fact that MSG is an amalgamation of dialects when it comes to phonetics and the decisions made when it comes to including or excluding sounds in the phonemic inventory. MSG pronunciation did not arise naturally but is a synthetic construct, largely based on *Bühnendeutsch*, a concept established in the 19th century (c.f. Elmiger 2019, Tkaczyk 2017) and the 1962 *Aussprachewörterbuch* by the Duden Verlag still stuck closely to the pronunciation norms described in *Bühnendeutsch* (Elekfi 1972). While the 2015 version of the Duden still references Sieb, the pronunciation norms suggested are based on the work of Mangold (2015) and the *Institut für Deutsche Sprache Mannheim* (Institute for the German language Mannheim) (Dudenredaktion 2015). This development is discussed in more detail below. Therefore the decision of which phonemes to include or exclude from the inventory of German is not straightforward. For now, I will give a brief overview of the vowel development in the history of the German language and show how this results in a synthetic concept of Standard German and why in conclusion this dissertation will only focus on one dialect of German, namely Westphalian.

MSG is a written language, which when spoken, is still colored by the speaker's own dialect pronunciation and even dialect-specific lexemes. Modern German dialects can be grouped into three broader areas: Low, Central, and Upper. The Upper dialects are considered those who have fully completed the second sound shift while Central and Low German dialects have only partially completed this sound shift. Within these three broader categories we find a variety of more finely grained subgroups. The dialect

observed in this dissertation is a Low German dialect, Westphalian, and can be seen in dark green in Figure 1 below.



Dialektale Großgruppen in Nordrhein-Westfalen und den angrenzenden Gebieten (Einteilung der Gruppen nach den Haupt-Isoglossen)

Fränkische Dialekte		Niedersächsische Dialekte
A. Niederfränkisch	B. Mittelfränkische Dialekte	Westniederdeutsch
Nordniederfränkisch	Ripuarisch	Nordniederdeutsch
Ostbergisch	Moselfränkisch	Westfälisch
Südniederfränkisch	Hessisch	Ostfälisch
	Pfälzisch	

Fig. 1: Dialect groups in North Rhine-Westphalia (Et Mikkel, https://en.wikipedia.org/wiki/File:Dialekte_in_Nordrhein-Westfalen.PNG)

Historically, German can be split into the periods of Old High German (OHG) (750-1050), Middle High German (MHG) (1050-1350), Early New High German (ENHG)

(1350-1700), New High German (1700-now), or Modern Standard German (MSG), all marked by sound changes of both consonants and vowels.

Splitting off from Indo-European, Old High German was marked by some simplifications in both phonetics and morphology, such as a reduction in declension and conjugation and the first and second sound-shifts in consonants as well as vowel changes, such as unaccented vowels, modified vowels (e.g. umlaut, a sound change in which a vowel is produced more like a following vowel) and vowel mergers (c.f. Chamber and Wilkie 2014, Rauch 2017, King 1965). When attempting to understand the German vowel system today, it is important to have an understanding of how vowels gradually changed over time. Looking at the vowel changes mentioned in the following from a phonological perspective, allows us to understand the differences and similarities present in modern German.

Vowel length, or duration, can be traced all the way back to Germanic, the stage that preceded OHG. When Germanic split off from Indo-European, the original set of long and short <a> and <o> underwent a vowel merger. Examples for this merger are shown in Table 1 and already show vowel length.

Table 1: Examples of vowel <a> <o> vowel merger (Salmons 2012:57)

Merger pattern	Indo-European	Germanic	MSG
*a = *a	<i>ghans</i>	<i>gans</i>	<i>Gans</i>
*o > *a	<i>orbho</i>	<i>arbi</i>	<i>Erbe</i>
*ā > *ō	<i>bhrātēr</i>	<i>brōpar</i>	<i>Bruder</i>
*ō = *ō	<i>plō</i>	<i>flōdus</i>	<i>Flut</i>

The short vowel <o> merged into short <a> in Germanic, and the long vowel <a> merged into the long vowel <ō> . This left a gap in the vowel system since no long <ā> was present in Germanic.

The shift from Germanic to OHG filled this gap with long <ā> occurring in environments with following *-xt (Salmons 2012), showing early attempts to keep the vowel system balanced with long-short pairs for each vowel. Additionally, for OHG, we see two main dialectal areas: High German (HG), as spoken in the South, and Low German (LG), as spoken in the North. Gloning and Young (2003) illustrate the differences in Old High German dialects with examples taken from *The Lord's Prayer* and claim that vowel differences were one major dialect marker between HG, and LG. These differences can still be observed in MSG, which is why this dissertation only focuses on one dialectal area of Germany as mentioned above. The vowel shifts occurring from Indo-European to OHG were monophthongization and diphthongization. While the PG diphthong <ai> becomes <ei> or <e> in LG, it becomes <e> or <á> in HG and the diphthong <au> becomes <ou> or <o> in HG but <ea> or <o> in LG. As for monophthongization, PG <e> stays <e> in LG but changes to either <ia> or <ie> in HG and PG <o> does not change in LG but changes to either <uo> or <ua> in HG. Examples to illustrate these changes are presented in Table 2 below.

Table 2: <ai> - <e> shift HG/LG (Szulc 1987)

Indo-European	OHG	English
<i>laizjan</i>	<i>leren</i>	<i>teach</i>
<i>saiwaz</i>	<i>seo</i>	<i>see</i>
<i>dailiz</i>	<i>Teil</i>	<i>part</i>

Table 2: <au> - <o> shift LG

Indo-European	OHG	English
<i>raudaz</i>	<i>rot</i>	<i>red</i>
<i>kaus</i>	<i>kos</i>	<i>choose</i>
<i>skaunaz</i>	<i>skon</i>	<i>beautiful</i>

These first two examples are examples of monophthongization in OHG, in which <ai> and <au> were replaced with either <e> or <o>, depending on the dialect area.

Table 3: <o> - <uo> shift HG

Indo-European	OHG	English
<i>moder</i>	<i>muoter</i>	<i>mother</i>
<i>fot</i>	<i>fuoƷ</i>	<i>foot</i>
<i>stol</i>	<i>stuol</i>	<i>chair</i>

This is an example of diphthongization in OHG, in which <o> shifts to <uo>. All examples show dialect dependent vowel shifts, whose effects can still be observed in MSG. The phrase “*liebe gute Brüder*” is an example of how LG and HG still differ in MSG based on these earlier occurring vowel changes. While in LG the phrase includes only monophthongs, in HG the phrase includes only diphthongs, which is preserved respectively in Westphalian and Bavarian German today.

Willmanns (1911:265) writes about the OHG vowel shifts:

“In demselben Maß als *ai* sich dem *e*, *au* sich dem *o* näherte, entfernten sich die alten *e* und *o* von ihrer ursprünglichen Form; *ai* und *au* wurden zunächst zu *ae* und *ao*, *e* und *o* umgekehrt zu *ea* und *oa*. Ein bewußtes Streben, die verschiedenen Laute auseinander zu halten, das sich in den Konsonantenverschiebungen bekundete, äußert sich auch hier.” [In the same way that *ai* became *e* and *au* became *o*, the former versions of *e* and *o* changed as well; *ai* and *au* became *ae* and *ao*, *e* and *o* became *ea* and *oa*. We can see the same conscious aspirations to keep the sounds apart, just like we see with the consonant shifts.]

He parallels the vowel shift to the consonant shift and attributes both to a theory of maximum dispersion. Maximum dispersion was first introduced by Liljencrants and Lindblom (1972) and states that vowels in a given system tend to be evenly dispersed throughout the vowel space, which is theorized to minimize confusion with other phonetic categories. So German using both duration and spectral cues to disambiguate vowels could be a strategy to maximally disperse the different categories.

While first mentioned by the name *theotisce* by papal legate George of Ostia (Keller 1978), German is first mentioned as a language by a German when Charlemagne decides that Latin literature should be translated into the language of the people (*tam latine quam diutiske = in Latin as in German*) (c.f. Chambers and Wilkie 2014, Horan, Langer, Watts 2009, McDonald 1972). This is important because the extant texts allow us to investigate the phonology of Old High German (OHG) and its dialectal differences with diachronic methods (Penzl 1971). The oldest known record of written OHG is *Der Abrogans*, a Latin to German dictionary (Baesecke 1930). Even in OHG, significant differences in the vowel systems of major dialect areas can be traced based on textual evidence. Some of these dialects are Franconian, Alemannic, and Bavarian and texts date back to the time from 750 to 1050 (Wiese 1987, Keller 1978). This fragmentation is largely due to the tribal nature of the Germanic peoples and Modern Standard German (MSG) still preserves these older forms of tribal languages to some extent in the form of the various dialects, Bavarian for example preserves some older diphthongs or monophthongs as is shown in the table 4 below and mentioned above in the discussion of LG and HG:

Table 4: Example of Bavarian diphthongs and monophthongs (Schirmunski 1962)

MSG	Bavarian
<i>ein</i>	<i>oan</i>
<i>Häuser</i>	<i>Haiza</i>
<i>Bäume</i>	<i>Bam</i>
<i>lieb</i>	<i>liab</i>
<i>gut</i>	<i>guat</i>

Penzl (1971) examines five different OHG works (Exhortatio ad plebem christianam, Isidor, Otfrid, Notkers, Otlohs Gebet) and notes that vowel quantity is present in all of these and expressed through different graphemic representations. Even in OHG, quantity is an important distinction in the vowel system.

In a recent history of the German language, Salmons (2012) mentions three major sound changes happening between OHG and Middle High German (MHG): weakening of vowels in unstressed syllables, umlaut, and a consonant change. While umlaut first occurred around 750 in the Rheinisch Franconian dialect (c.f. Iverson et al. 1994, Kyles 1967, Voyles 2011), it spread to other dialects of German as well. In Rheinisch Franconian umlaut first only affected <a>, but by the eleventh century umlaut had spread to <o> and <u> as well as <ou> and <uo>. Umlaut caused what were likely back vowels to become fronted and was not evenly spread throughout all dialects (c.f. Voyles 1976, Penzl 1949, Twaddell 1938). Additionally, weakening of unstressed vowels can be traced by comparing OHG forms to MHG forms, such as *OHG suntia* vs *MHG sünde*, in which the word-final <ia> turns into schwa (Salmons 2012). This important sound change process created a series of new phonemes, which are still present in Modern Standard German (Penzl 1949). Another modern distinction that we can

observe in older stages of the language is the quantity distinction in vowels. MHG had 23 stressed vowels (Szulc 1987), which show the long-short distinction we still see in Modern German today: the short vowels <i, e, ě, ä, a, ü, ö>, the long vowels <i, e, ä, a, iu, ö, u, o> and the diphthongs <ie, ei, üe, öü, uo, ou>.

The sound changes occurring from MHG vowels to Early New High German (ENHG) vowels are monophthongization, diphthongization, open syllable lengthening, and closed syllable shortening (Salmons 2012). The old diphthongs turn into long monophthongs in (E)NHG and we still see them as long vowels in MSG today. Table 5 shows vowel length in MHG and (E)NHG and word examples.

Table 5: Examples of length in MHG and (E)NHG (Salmons 2012:232)

MHG	(E)NHG	Word examples
<i>ie</i>	<i>i:</i>	<i>liep</i> > [li:p]
<i>uo</i>	<i>u:</i>	<i>bruoder</i> > <i>bruder</i>
<i>üe</i>	<i>ü: [y:]</i>	<i>müede</i> > <i>müde</i>

Additionally, there was lowering of the short vowels and raising of the long vowels (Salmons 20212), which is likely why the tense/lax distinction was raised for MSG.

Having described some important sound changes in the phonological features of OHG and MHG, the literature cited above has effectively set up the context of looking at MSG in terms of its development and evolution of the language, especially with regard to duration.

1.2.2 Quantity in German - tense/lax or long/short?

Quantity in German vowels has been described along two dimensions, either as tense/lax or as long/short. The following section will give an overview of the literature on the German length distinction in vowels and why there was disagreement about which individual phonemes to include in the vowel system of Modern Standard German based on whether the phonological distinction between long and short vowels was based on tenseness or duration.

The development of both a graphemic and phonological union of German was mostly driven by the reformation, grammarians, teachers and writers, and the *Sprachgesellschaften* of the 19th century (c.f. Russ 2002, Berns 1988, Keller 1978). With the foundation of the German Empire in 1871, a norming of the language became more important to form a national consciousness, and the abolishment of serfdom at the beginning of the 19th century and the industrial revolution in the 1830s caused more people to move to the cities. This pushed the development of supra-regional vernaculars (Stedje 2007:172ff). The efforts of grammarians and writers were targeted toward uniformity, regularity, and clarity of the language (Keller 1978).

At the end of the 19th century, a conference in Berlin held by phoneticians reached some consensus about a standardized pronunciation of German. One of the participants was Theodor Siebs, a professor of linguistics (Waterman 1966:174, Keller 1978). His work is widely regarded as a base for Modern German standard pronunciation (c.f. Moulton 1962, Elmiger 2019, Tkaczyk 2017, Elekfi 1972). In Siebs' *Beratungen zur ausgleichenden Regelung der Bühnenaussprache* (1898) 22 vowel

phonemes are suggested, 19 monophthongs and three diphthongs: <l, ε, Y, ö, a, U, ɔ, i:, e:, ä:, y:, ø:, a:, u:, o:, ae, ɔø, ao>. This phoneme system is based on the opposition of tense and lax; generally tense vowels are produced with greater muscle tension than lax vowels (Delahunty and Garvey 2004). The tense lax distinction is often based on phonotactics and in languages like English and German, tense vowels can occur freely at the end of monosyllabic words while lax vowels mostly occur in monosyllabic words only if they end in a consonant. The duration patterns of tense and lax vowels are seen as secondary, but tense vowels are overall longer than the lax vowels of the same height (Nearey 2006). While the feature has been described as a degree of muscular tension (Jones 1918, 1964), Jakobson, Fant, and Halle (1952) define the tense-lax distinction as follows:

“In contradistinction, to the lax phonemes, the corresponding tense phonemes display a longer sound interval and larger energy (defined as the area under the envelope of the sound intensity curve) [...] In a tense vowel the sum of the deviation of its formants from the neutral position is greater than that of the corresponding lax vowel. (1952: 36) Production. Tense phonemes are articulated with greater distinctiveness and pressure than the corresponding lax phonemes. The muscular strain affects the tongue, the walls of the vocal tract, and the glottis. The higher tension is associated with greater deformation of the entire vocal tract from its neutral position. This is in agreement with the fact that tense phonemes have a longer duration than their lax counterparts. The acoustic effects due to the greater and less rigidity of the walls remain open to question.” (1952: 38)

While Siebs made an effort to promote a standardized German pronunciation, spoken German shows substantial variation and therefore several different phoneme inventories have been suggested for MSG in the literature. It is important to note that *Bühnenaussprache* was mostly intended to be used as a normalizing force used in educational contexts as well as on the stage, on radio, or television (Siebs 1969, 2020). Spoken German will always show traces of regional dialects, even if a speaker attempts to speak MSG.

Similar to Siebs, Szulc (1987) also suggests a vowel inventory largely based on the tense lax distinction, his phoneme inventory consists of the tense vowels: <i:, i, u:, u, y:, y, e:, e, o:, o, ø:, ø, a:, a> and the lax vowels: <ɪ, U, Y, ε, ə, ɔ, ö, a>.

The debate whether the distinction in the German vowel system is one of tense-lax or long-short influences the suggested vowel inventories. While Keller (1978:554) describes the German vowels along the dimension of tense and lax, he mentions that scholars are “not in agreement about the primary opposition: is it short - long or lax - tense?”.

Using the long-short distinction, Werner (1972) lists 15 monophthongs and three diphthongs. He suggests the eight long monophthongs /i:, y:, e:, ø:, a:, u:, o:, æ:/, and the seven short monophthongs /ɪ, ʏ, ø, a, o, u, e, a, æ/, as well as the three diphthongs /ai, oi, au/. While using the long-short distinction for monophthongs, he writes that it is unclear whether the difference is quantitative or qualitative and it is also unclear whether the diphthongs are indeed independent phonemes or merely a combination of two isolated monophthongs. Similarly, Wiese (2000:11) lists the following phonemes as vowels of German: /i:, ɪ, y:, ʏ, e:, ε, ε:, ø:, œ, a:, a, o:, ɔ, u:, ʊ, ə, ɐ, ai, aʊ, ɔʏ/. However,

instead of listing short vowels as phonemes, he lists long vowels as phonemes and their short counterparts as allophones (2000:18), which implies that instead of being different phonemes they are instead in complementary distribution. He suggests that for /e:/ vs /ɛ:/ and /a:/ vs /a/ duration is the only contrastive feature and that they do not show differences in quality (2000:22).

Vowel length has been discussed as having three levels in Menzerath (1939), Martens (1955), and Müller (1956, 1958), who suggest that in addition to the distinction between long and short vowels *überlang* (overlong) should be added as a third duration step. They try to illustrate these with minimal pairs as in *reißt - reist, fließt - fliehst, Rute - ruhte*. This suggestion has been rejected by later research (c.f. Hanhardt et al. 1965, Delack 1972, Newton 2019).

According to Handke (Campus, T.V.L. 2012), German has 16 monophthongs and 8 diphthongs. The seven long monophthongs are /i:, y:, e:, ø:, a:, u:, o:/, he adds that /ɛ:/ is often merged with /e:/. The nine short monophthongs are /ɪ, ʏ, ɛ, œ, a, ɔ, ʊ, ə, ɐ/ and the eight diphthongs are /ie, yə, eə, ai, ui, oə, ɔʏ, aʊ/.

While there is a debate on the tense-lax versus long-short distinction, this dissertation will use the terms long-short as referring to quantity for several reasons. For one, the tense and lax features are not present in the IPA notation (Durand 2005). Jones writes that “it is extremely difficult to determine in the case of the opener vowels whether the sensation of ‘tenseness’ is present or not” (1964: 39-40) and Lass has described the distinction as a “contentless dichotomizing operator” (Lass 1976: 9-10). Additionally, tense-lax and duration correlate with each other in that “lax vowels tend to be short whereas tense counterparts are long.” (Kwon 2011:606).

While the literature referenced above is making the decision of whether German long and short vowels differ in tenseness or duration based on theory, it is important to look at the evidence from production and perception to define what duration looks like acoustically and how spectral information ties into the distinction of German long and short vowels. The next chapter will give an overview of acoustic and perceptual evidence for quantity in German.

1.2.2.2 Using acoustic evidence to investigate quantity and quality

Evidence from studies using non-native vowels suggests that native German speakers are indeed sensitive to both quality and quantity (temporal and spectral information) (Bohn and Flege 1990, Escudero et al. 2009). But which cue is used primarily in the disambiguation of native German long-short vowel pairs and what are the defining acoustic features of quality and quantity in German? Recall that the focus of this dissertation is on vowels as acoustic events. Therefore, it is necessary to first establish the specific acoustic features that define quality and quantity.

Spectral information refers to the specific formant frequencies associated with a vowel. The different lip shapes used in the production of different vowels filter the frequency generated by the source, which causes the output waveform to have varying combinations of component frequencies and amplitudes, or a different spectrum, depending on the shape a speaker needs to make to produce a target sound. By reshaping the vocal tract, certain frequencies are amplified and others dampened out,

these amplified resonance frequencies are called formants. Different formants create different qualities of sound. For vowels, the resulting complex waveform is the sum of a set of sinusoid waves with different frequencies and when visualized, the frequency and amplitude result in a spectrum (Zsiga 2013:116). In vowel production, formants also correspond to the position of the tongue. The first formant (F1) corresponds to the degree of tongue height and the second formant (F2) refers to tongue frontness. The importance of the first two formants has been well studied (c.f. Peterson and Barney 1952, Fant 1959, Hillenbrand, Getty, Clark and Wheeler 1995) and therefore F1 and F2 have been said to provide sufficient acoustic information to identify a vowel (e.g. Delattre et al. 1952, Dunn 1950, Peterson and Barney 1952). By representing vowels in this two-dimensional space, a language-specific vowel system can be mapped and compared to other languages with different configurations in the F1/F2 plane. Plotted in the two-dimensional F1/F2 space, the German vowel system appears as shown in the Figure 2 below:

Figure 7: German vowels produced in /hVt/ syllables, presented in an F₁/F₂ space.

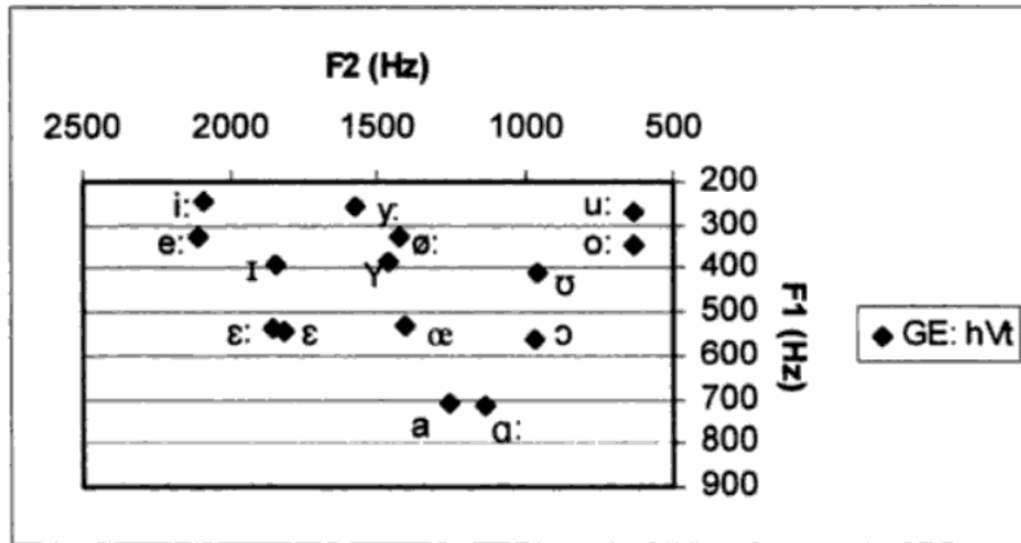


Fig. 2: German vowels produced in /hVt/ syllables (Steinlen 2005:79)

Especially for languages like German with large vowel systems as shown in figure GERSPA1, using only the first two formants is an insufficient way to characterize vowel categories and mark the contrast between categories which overlap in their formant frequencies (Fant 1960; Maurer et al. 1992). Other acoustic cues, such as duration (c.f. von Essen 1979, Heike 1969, 1970, 1972, Lindner 1976, Weiss 1976, Sendlmeier 1981), are used by listeners to disambiguate vowels. Duration corresponds to the quantity of the vowel or the length of a vowel.

Several studies have investigated whether quality or quantity is the primary cue in distinguishing between long and short vowels in German.¹ While some studies show that spectral characteristics are weighted heavier when identifying German long-short

¹ Depending on whether a cue is primary or secondary in a language, listeners assign different degrees of perceptual attention to an acoustic cue (c.f. Scobbie 1998, Escudero 2000, Nittrouer 2000).

vowels (c.f. Bennett 1968, Ungeheuer 1969, Strange and Bohn 1998), other studies show the opposite and claim that temporal information, specifically vowel duration, is the main cue in distinguishing between the long-short vowel pairs (c.f. von Essen 1979, Heike 1969, 1970, 1972, Lindner 1976, Weiss 1976, Sendlmeier 1981, Bennett 1968). There is no clear consensus in the literature, and most studies do not investigate the role of secondary cues. For example, Riad (1995), Wiese (1996), and Lahiri and Dresher (1999) all treat vowel length in German as a purely prosodic phenomenon and therefore argue that German does not show vowel quantity as a discriminative feature at all, while on the other hand, several studies (c.f. von Essen 1979, Heike 1969, 1970, 1972, Lindner 1976, Weiss 1976, Sendlmeier 1981, Bennett 1968) have shown that duration alone was a sufficient cue to distinguish between vowel categories for German listeners.

Investigating the role of these cues acoustically is important to gain a detailed picture of the interplay between quality and quantity in German. Several studies have focused specifically on the acoustic cues corresponding to quality and quantity to tease apart cue usage in German. For example, Strange and Bohn (1998) used manipulated tokens, showing that when duration information was removed from the vowels, identification errors increased significantly. Similarly, Tomaschek et al. (2014) have found duration to be the main cue when discriminating between long and short /u/-/o/ vowel pairs, and Tomaschek et al. (2011) have shown that vowel length in German even matches the criteria for categorical perception.² While duration was found to be the primary cue to distinguish between long and short vowels, Tomaschek (2013) has also found that quality serves as a secondary cue.

² Even though this has been shown mainly for consonant perception.

In contrast to the studies mentioned above, the use of spectral information in the perception of German long-short vowels has been extensively shown as well (c.f. Bennett 1968, Ungeheuer 1969, Strange and Bohn 1998). Using acoustically manipulated tokens, Heike (1969) showed that when manipulated only for duration, only the /a:/-/a/ distinction showed a clear boundary shift between the long and the short vowel.³ Her results show vowel-specific patterns for other German vowels with /o:/ being identified as /ʊ/, and /e:/ as /i/, when duration was manipulated. If duration was the only cue used in the disambiguation of German long and short vowel pairs, these vowels should have been identified as their short counterpart. Heike's findings, therefore, show that there is a complex relationship between duration and spectral cues. Similar complex patterns have been found by Bennett (1968), who compared the use of duration in the perception of both English and German vowels using synthetic vowels manipulated for both duration and spectral information by combining 4 spectral values with 4 durations. The results show that for both languages spectral cues were used primarily and duration was used secondarily. However, for German, the results showed vowel-specific patterns: For example, duration had a greater effect in the /ʊ/-/o:/ pair, while spectral information was used primarily in the /ɔ/-/o:/ pair. Weiss (1976) also found vowel-specific patterns. His results show that the importance of spectral information versus duration information was a function of vowel height with listeners relying more on duration to identify low vowels, but using spectral information to identify high vowels. These results have been replicated by Sendlmeier (1981) as well.

³ The /a:/-/a:/ distinction seems to be a special case in the German vowel system and studies have shown them to be identical in quality (Duden 1990, Siebs 1969, Ungeheuer 1969, Kohler 1990, Martens and Martens 1961, Hoffmann 2011, Sendlmeier & Seebode 2006).

Taken together, this shows that German listeners use a combination of quality and quantity when disambiguation between long and short vowels.⁴ Production studies also show long and short vowels differing along multiple acoustics dimensions. Pätzold and Simpson (1997) worked with a corpus of 22 speakers (11 male, 11 female) from the North of Germany. The data came from the Phon-Dat90 corpus (Thon and Dommelen 1992) collected and labeled at the IPDS Kiel (IPDS 1994) containing all vowels of MSG, in which two sets of 100 sentences were read. The sentences contain an average of five words per sentence. The sixteen monophthongs and three diphthongs were analyzed using the first three formants. They found that the short vowels are all more central than their long counterparts in the F1/F2 space and that only /a/ and /a:/ are similar in spectral quality, using duration as the main cue for distinction, whereas the other long-short vowel pairs show clear spectral differences. These findings have been replicated by Steinlen (2005) and Jørgensen (1969) who show that all tense and lax vowel pairs show significant spectral differences and that the lax vowels are located more centrally in the F1/F2 plane. Heid, Wesenick, and Draxler (1995) found similar patterns when looking at the overall duration of long and short vowels in German. They found that while phonologically long vowels are on average longer than the short vowels (97ms long vowels, 64ms short vowels), the vowels show systematic spectral differences depending on whether they are long or short: “although the overall distribution is fairly similar, it appears [...] that long vowels concentrate mainly in three regions: back/round/high, front/high, central/low. The short vowels [...] are distributed

⁴ However, the long-short vowel pairs he used for the German listeners were chosen based on the author's personal judgment of which vowels sound similar and not determined based on the actual spectral qualities of the vowels as observed in production.

more regularly with a higher concentration in the center of the vowel space.” (Heid et al. 1995:4).

Taken together, the question of whether German listeners rely primarily on quality or quantity differences is unclear with studies finding conflicting evidence. This dissertation seeks to provide a clear answer to whether quality or quantity is used as a primary cue in vowel perception as well as the interaction of these cues with secondary cues. The following section will provide an overview of the quality-quantity debate in the perception of German vowels.

1.2.3 The quantity-quality debate

As discussed above, German listeners seem to use both quality and quantity differences in the perception of long-short vowel pairs. However, there is little consensus on which cue is used primarily as shown in the discussion of the literature above and research stating that either only quality or only quantity is distinctive while the other one is redundant, ignoring the use of secondary cues (c.f. Riad 1995, Wiese 1996, Lahiri and Dresher 1999, Vennemann 2000, Mangold 1990, Delattre 1969, Vernon 1976, Maack 1951, 1954, Jessen 1993, Weiss 1977).

What defines a primary feature? The functional-structural framework (Jakobson, Fant & Halle 1952, Jakobson & Halle 1968) states that a primary or distinctive feature is invariant across contextual variation. If quality is the non-redundant feature in German, it should stay invariant across different contexts, such as stressed vs unstressed

positions. Ramers (1988) hypothesizes that there are two different ways in which quality and quantity can be distinguished in terms of primary and secondary cues. The first hypothesis states that vowels that aren't stressed tend to reduce differences in quantity more readily than in quality. This would mean that quality differences are more robust and stable. In unstressed syllables, vowels would only be distinguished in quality and not in quantity making quality the invariant and primary cue. Using this framework, it could be determined from production alone whether quantity or quality is the primary feature in the distinction of German vowel pairs. Mangold (1990) has shown neutralization of quantity differences in long-short vowel pairs in positions before main stress. Similar results have been shown by Delattre (1969) and Vernon (1976). These studies support the use of quality as a primary cue. However, results showing no significant reduction of quantity differences in stressed vs unstressed positions can be found in Maack (1951, 1954), which supports the use of quantity as a primary cue. The existence of minimal pairs for long and short vowels is another indicator of contrastive quantity in German, as minimal pairs are traditionally used to establish phonemic status in a language (c.f. Brown 1995, Levis and Cortes 2008, Maye and Gerken 2000). Figure 3 shows minimal pairs for all German long-short vowel pairs:

Minimal pairs involving vowel quantity (Wiese 1996: 11)

<i>bieten</i> ‘offer’ ~ <i>bitten</i> ‘request’	[i:] ~ [ɪ]
<i>Hüte</i> ‘hats’ ~ <i>Hütte</i> ‘hut’	[y:] ~ [ʏ]
<i>Beeten</i> ‘(flower) bed’ ~ <i>Bett</i> ‘bed’	[e:] ~ [ɛ]
<i>sehen</i> ‘see’ ~ <i>säen</i> ‘sow’	[e:] ~ [ɛ:]
<i>äße</i> ‘eat (1sg.subj.)’ ~ <i>esse</i> (1sg.ind.)	[ɛ:] ~ [ɛ]
<i>Höhle</i> ‘cave’ ~ <i>Hölle</i> ‘hell’	[ø:] ~ [œ]
<i>Schal</i> ‘scarf’ ~ <i>Schall</i> ‘sound’	[a:] ~ [a]
<i>Ofen</i> ‘oven’ ~ <i>offen</i> ‘open’	[o:] ~ [ɔ]
<i>spuken</i> ‘spook’ ~ <i>spucken</i> ‘spit’	[u:] ~ [ʊ]

Fig. 3: German minimal pairs for long-short vowels taken from Wiese (1996)

Keller (1978:554) suggests that vowel duration is a prominent phonetic feature of German and advises against relegating vowel length to a subsidiary position.

A third approach could be seen as an alternative hypothesis to the use of either quality or quantity as a primary cue and instead suggests that both quality and quantity are reduced to the same degree in unstressed syllables and therefore no vowel is categorized solely by one or the other. Rather, both are equally important in the perception of long and short vowels. For example, Jessen (1993) has shown both quality and quantity to be invariant in his experiments and argues that neither quality nor quantity can be the sole primary cue. Furthermore, his results show vowel-specific patterns: quality is invariant in different stress positions for all but non-low vowels. He argues that quality is the primary cue for non-low vowels in distinguishing between long and short, while quantity is distinctive for low vowels. This vowel-pattern-specific approach has also been suggested by Weiss (1977), who states that high vowels are

distinguished primarily by quality, low vowels by quantity, and mid vowels by a combination of both.

While some of the literature above suggests a complex relationship between quantity and quality in the perception of German long-short vowels, the exact nature of this relationship is still unclear, as well as whether quantity or quality is the primary cue used in the distinction. The experimental evidence cited in this dissertation provides evidence for both theories, therefore still leaving the question of which cue is used primarily unanswered. This dissertation investigates the use of duration and spectral cues, as well as the use of VISC in the classification of long-short vowel pairs by native German speakers. The aim of the experiments conducted is to achieve a better understanding of the relationship between quantity and quality in German vowels.

Revisiting claims made about the production and perception of vowel quantity and quality in German, previous literature has stated that either duration or spectral cues are the sole perceptual cue used by listeners to differentiate between minimal pairs such as *Kamm* and *kam*. In the first experiment, a production study will answer the question of how German vowels are produced with respect to acoustic cues, such as spectral features, vowel duration, fundamental frequency, and F1, F2, and F3 at onset, midpoint, and offset of the vowel and whether there are significant differences in these properties between long and short vowels. Additionally, this study will investigate cross-language vowel perception to test whether the German quantity/quality differences are perceptually salient enough to elicit a phonemic contrast in non-native speakers, that is to not be collapsed into the same vowel category in English. Listeners should be more sensitive to the cue that is more salient and use it to disambiguate

vowel instances, so if quantity is the more salient cue German /a:/ and /a/ should not be collapsed into the same English category. Of course, cue weighting from English could obscure the salience of German cues.

Experiments two and three will investigate which acoustic cues listeners utilize in the perception of long-short vowels. Experiment two uses a synthetic 5-step continuum of manipulated vowel durations, while experiment three uses a synthetic 5-step continuum of manipulated vowel formants (F1-F3). Taken together, the experiments will address the question of whether duration or spectral information is the only perceptual cue used by listeners, or if both cues are used in conjunction and how they are weighted.

Lastly, experiment 4 will investigate the use of spectral information in more detail by looking at the importance of VISC by German listeners using resynthesized silent-center and silent-onset/-offset vowels. If German listeners rely mostly on spectral cues from the center of the vowel, the silent-center vowels should be identified with lower accuracies than the silent-onset/-offset vowels.

In order to investigate the importance of different acoustic cues in speech perception, a theoretical basis of perception must be established first. The next section will give an overview of different theories of speech perception. The question of how listeners perceive speech and which elements of the signal listeners are sensitive to when the signal itself is variable is a longstanding debate in linguistics. The next section describes five major theories of speech perception and the shortcomings of each.

1.3 Theories of Speech Perception

How do listeners extract spectral and temporal information from the speech signal? This is a complex question when considering that there is substantial inter- and intra-speaker variability and the fact that the speech signal is continuous, even though the individual units are discrete. Cues are therefore transmitted quickly and simultaneously. There has been a focus on the mappings between properties found in the speech signal and units such as phonemes. However, these mappings are complex and there is still no complete explanation of how humans recognize vowel and consonant sounds. In the following section, an overview of the major theories of speech perception is given.

Phonetic segments often display different acoustic information, like varying formant frequencies and durations, and researchers have been trying to explain how humans are able to deal with this lack of invariance while still mapping acoustic inputs to phonemic targets. The invariance problem is concerned with the fact that listeners have to deal with variable input to a constant feature or phonetic category (c.f. Appelbaum 1996, Fowler and Magnuson 2012, Blumstein 2021). There is no one-to-one mapping between linguistic units and the physical objects, which show extensive variation. Nonetheless, listeners have no problem decoding the speech signal. Variation is present within speakers and between speakers, based on different contexts and styles. Speakers have accents and speak a dialect and they can also hyper- or hypoarticulate, or coarticulate. Additionally, speech can be influenced by other conditions like lack of sleep, injuries of the tongue, height, gender, age, speech impairments, and environmental conditions like noise, social setting, or register.

The different theories in speech perception use different approaches to answer the questions of which units are perceived and which processes are utilized in the process of perceiving speech. McQueen (2005:265) summarizes:

“One recurring issue in this debate has been whether the objects of speech perception are fundamentally acoustic in nature (Kluender, 1994), or are gestural in nature (Fowler, 1986, 1996; Liberman & Mattingly, 1985), or are the product of pattern-recognition processes (Massaro, 1987; Nearey, 1997). Another related issue has been whether speech perception calls on special processes (Liberman & Mattingly, 1985), or depends on general auditory processes that are also used in the perception of other complex sounds (Pastore, 1981; see, e.g. the debate between Fowler, Brown, & Mann, 2000, and Lotto & Kluender, 1998, and Lotto, Kluender, & Holt, 1997).”

There are five major theories of speech perception trying to answer these questions: articulatory phonetic theories, gestural theories, segmental theories, auditory/acoustic theories, and probabilistic models. Articulatory phonetic theories are production-based while auditory phonetic theories are perception-based. Motor theories state that perception is mainly related to production and the discrete units are gestures and that speech perception is special and unique from other perceptions. Segmental theories like the LAFS or TRACE state that a sequence of transformations from sounds to objects governs speech perception. Auditory theories posit that speech perception is derived from general properties of the auditory system and is not operating with a specific speech perception module. Probabilistic theories are non-analytic and state that information about individual instances of speech is stored as episodic information. The following will give a brief description of the different models in the five categories.

1.3.1 Gestural theories

In 1950, Alvin Liberman, Franklin Cooper, and Pierre Delattre developed the motor theory of speech (Delattre et al. 1951, 1952, 1955, 1964; Liberman 1957; Liberman et al. 1952, 1954, 1956), which states that speech is perceived by the listener based on identifying vocal tract gestures instead of invariant acoustic patterns. The theory originally included a specific decoder or speech module, which maps gestural patterns as the objects of perception. Speech perception is based on speech production and listeners use their own knowledge of how to produce a sound as the template to decode the speech signal. While this theory has undergone significant changes since its earliest version, at its core it still claims that articulatory events are the base of speech perception (Diehl, Lotto, and Holt 2003). This also solves the invariance problem, as the neuromotor commands to the articulators, or intended gestures, are relatively invariant (Liberman et al. 1967). Some evidence in support of this theory can be observed in the McGurk effect, in which listeners are presented with an audio stimulus, presenting the acoustic information, and a visual stimulus, showing the speech gesture. Experiments with conflicting stimuli have shown that speech perception relies on an audio-visual integration effect, which causes listeners to mishear speech if the shown gesture and the acoustic cues presented are different (McGurk and McDonald 1976). When presented with conflicting audio and visual stimuli, the speech perception is influenced by the shown production gesture. Another effect that has been used to support the gestural claim is the complications that can arise when listeners are presented with acoustic stimuli only and no supporting visual clues are presented to them, as in

telephone conversations. Some brain studies have also shown that when being presented with acoustic stimuli, vocal tract muscles, the motor cortex, and the premotor cortex are activated as well as mirror neurons⁵ (Lotto et al. 2009, Hickok 2010, Rogalski et al. 2011). However, the motor theory of speech ignores the acoustics and just looks at gestures, which are described to be invariants, and therefore the problem of acoustic variance needs not to be dealt with. However, gestures are subject to variation as well, based on coarticulation and speaking style (Mattingly 2019). The motor theory does not address this. While listeners are able to identify sounds in isolation fairly well, speech perception is also affected by contextual clues. Multiple stimulus sources are involved in perceptual processes. Another factor arguing against the motor theory is that listeners are able to identify sounds that they are not able to produce, such as infants that perceive the sounds of their language well before they can produce them (Bruderer et al. 2015). Infant studies with pacifier sucking rates have shown that arousal goes up when presented with phonemes of their L1 versus unfamiliar sounds (Bruner 1973, Jusczyk et al. 1990, Barca 2019). Additionally, acoustic perception extends beyond human speech perception, with studies showing that animal sounds and even ambient sounds like door slams are just as readily perceived and humans are able to tell the size of an object from their resonances, e.g., how big an object is (c.f. Carello et al. 2005, Bleak and O'Meara 2013, Zsiga 2013).

Another gestural theory is Fowler's direct realist approach (Fowler 1981, 1984, 1986, 1989, 1994, 1996) which states that perception recovers the sound objects directly and without any form of mediation. Like in motor theory, listeners perceive

⁵ Whose effects and existence are still controversial in psychology.

gestures, not phonemes, based on the acoustic information. Fowler compares speech perception to general perception, such as visual perception, and argues:

“Perceptual systems have a universal function. They constitute the sole means by which animals can know their niches. Moreover, they appear to serve this function in one way: They use structure in the media that has been lawfully caused by events in the environment as information for the events. Even though it is the structure in media (light for vision, skin for touch, air for hearing) that sense organs transduce, it is not the structure in those media that animals perceive. Rather, essentially for their survival, they perceive the components of their niche that caused the structure.” (Fowler 1996:1732)

Thus, the listener is able to recover the physical properties of the gesture through the rich acoustic signal and no symbolic representation in the mind is needed. Similarly, indirect realism states that human perception does not correspond to the reality of objects, but is mediated by human concepts and mental representations (Perkins 1983). It is suggested that perception is not governed by information processing and data flow, but by concepts that restructure data and map it to abstract units in the brain. This means that no matter what the physical stimulus is, when the same neural path is activated, the same perceptual experience will arise. In this way, acoustic variance is dealt with since the object of perception is not the actual physical stimulus but the concept, or underlying form, that is triggered by it. Allophones are merely physical objects mediated by a concept in the mind and then mapped to an abstract unit of perception, effectively by-passing the invariance problem.

1.3.2 Segmental theories

In comparison, segmental theories like TRACE (McClelland and Elman 1986) and LAFS (Klatt 1979) suggest that a strong link between production and perception, as suggested in motor theory, is unnecessary. Coding of the acoustic signal is based on auditory processes that use a form of intermediate representation and an information processing framework. In 1979 Klatt suggested the LAFS (lexical access from spectra) model, which combined phonological and acoustic properties into a spectral sequence decoding network structure. The LAFS was one of the first automatic speech perception models. His model is not a model of single phoneme recognition, but a model of word recognition based on sequences of spectral templates. These templates are context-sensitive, since they categorize the acoustics of phonemes in different phonetic environments by encoding spectral characteristics of individual segments and their transitions from one segment to the next, almost like an n-gram model. The spectra are computed every 10ms and then compared to the spectral templates stored in the network. Klatt's model is auditory based, using acoustic cues from the input signal to decode the speech stream in real time. His network is a finite-state machine, one specific path can only result in one specific output. Phonetic perception is mediated by the lexical network. Similarly, the TRACE model uses three levels representing features, phonemes, and words. Each level is interconnected and connections can be excitatory when levels share common properties, or inhibitory when one feature is activated and other features are therefore inhibited based on the phonotactic constraints of a language (Wright et al. 1999). One problem with these models mentioned by Wright et.

al (1999) is that they are unable to account for lawful acoustic variation due to prosody, rate, or speaker differences. Another issue that has been brought up is that it is neurally and cognitively implausible to have network access to multiple instantiations at the same time (Cutler 1995, McClelland & Elman, 1986).

1.3.3 Auditory/acoustic theories

Auditory, or acoustic, theories such as Nearey's pattern recognition model (1995) suggest that speech recognition is essentially acoustic pattern recognition. In his model, the real-time objects of perception are well defined auditory patterns that listeners use to decode the speech signal. He calls his theory a double-weak approach: speech cues are directly mapped onto phonological units no bigger than phoneme size. Gestures and sounds are linked only indirectly through separate links to shared symbols. There is no simple relation of either articulation or acoustics to the phoneme. Perception and production are distinct but cooperative systems and listeners integrate context sensitive information. Listeners are responding to the distinct patterns of the speech wave and match the incoming auditory patterns to their stored patterns to identify sounds. Phonetic constancy is reached by processes of normalization. Nearey developed a sliding template model to handle non-invariance found in the acoustic signal, where formants are normalized in a log-space with constant ratios (Nearey 1989).

1.3.4 Exemplar theories

Probabilistic models like exemplar models (Pierrehumbert 1998, Hintzman 1986, Hay, Warren, and Drager 2006) are non-analytic and state that information about individual instances of speech are stored as episodic information. Mental representations are not highly abstract or redundant. The episodic traces are linked on multiple levels and form multi-dimensional representation clouds, with clouds being constantly updated and even reshaped based on new examples. Categorization depends on the denseness of the cloud and a cloud having a center of density. New stimuli are compared to the clouds and the more similar they are to the center of the cloud, the faster and more accurate the categorization will be (Zsiga 2013:194). So, when listeners hear a token of a phoneme, for example /a/, exemplars are stored in detailed memory instead of an abstract generalization. These detailed traces are used for future judgments using a similarity function to measure the distance between the observed token and the stored tokens. Those models can be interpreted as a form of Monte Carlo approximation since they approximate an expectation of which category an exemplar belongs to. Monte Carlo methods work by repeated random sampling to estimate a probability distribution on the base of which the function expectation is calculated (c.f. Karp et al. 1989, Dagum et al. 2000, Shi et al. 2008, Giles et al. 2015). More sophisticated and refined versions of this theory have used Bayesian inference and can account for human performance even with only a few exemplars available. Studies (Shi et al. 2008, Shi et al. 2010) using simulations have shown that exemplar models, when implemented with Bayesian

inference⁶, are able to provide a mechanism of perception that is flexible and does not rely on any abstract representations of real-world data.

1.4 Normalization and the invariance problem

Acoustic theories have long tried to account for the invariance problem as acoustic cues are so vastly different between and even within speakers. Processes of normalization have tried to account for how listeners are able to map acoustic variants to one phonemic category, despite the ambiguity in the signal. Barreda (2020:2) defines vowel normalization as “the perceptual process that determines vowel quality from speech acoustics, including the ability to associate acoustically dissimilar vowels with a similar perceived vowel quality.” Ladefoged and Broadbent (1957) have argued that three different types of information are conveyed when speakers produce vowel sounds: phonemic information (vowel identity), anatomical information (vocal tract information), and sociolinguistic information. All of these characteristics can influence the acoustics of a vowel, such as formant frequencies (Peterson and Barney 1952, Labov 2001).⁷ The different theories of normalization have focused either on vocal tract normalization (Nearey 1978) focusing on steady-state vowels only and therefore not accounting for how different talkers use different cues or cue combinations, or how coarticulation influences acoustics (Dorman, Studdert-Kennedy, & Raphael, 1977) and how listeners

⁶ Which refers to using Bayes theorem to deduce properties about a probability distribution from the data.(see Dempster 1968, Box and Tiao 2011)

⁷ Both anatomical information and sociolinguistic information have been treated as noise in the past and therefore as information that should be removed (Adank et al. 2004).

compensate for that (Johnson, 1991; Nusbaum & Morin, 1992), or talker normalization (Lobanov 1971), using various sources of information about the given acoustics and talker which can be derived either from context or from the utterance itself (Ainsworth, 1975). Normalization methods deal either with phoneme extrinsic information or with phoneme intrinsic information.⁸ Intrinsic methods use nonlinear transformations of the frequency scale (log, mel, bark) and transformations based on a combination of formant frequencies (Adank et al. 2004). These methods usually preserve more information and are trying to model human speech perception since listeners are flexible in their perception and use many context clues, like vocal tract estimation, to account for variant inputs (Nusbaum and Magnuson 1997). Extrinsic methods on the other hand employ formant information from the point vowels of a talker and remove more talker information. These methods are often employed in automatic speech recognition, as they yield more correct identification rates for between-talker scenarios (Adank et al. 2004). More detail on the different methods will be given in the following section.

1.4.1 Extrinsic theories

Typically, extrinsic theories posit that in order to normalize, or know that different formants are still mapped to the same vowel category, a listener needs acoustic information that consists of more than one phoneme token of a talker. Based on multiple phoneme tokens, listeners develop a talker-specific coordinate system in order to

⁸ They can be further categorized into active (open-loop) or passive (closed-loop) theories (Barreda 2020). While the passive theories suggest a static relationship between acoustic cues and perceptual interpretation, active theories state that inputs are monitored and modified according to the context.

normalize and account for speaker-specific vocal tract information. These theories often also address issues of speech perception on a broader scale. Utilizing contextual information like speaker size or gender to establish a reference system and know what kind of formant changes should not change vowel quality was proposed in Nearey's uniform scaling method. Uniform scaling uses a speaker-dependent scaling term (Nearey 1978). By accounting for vocal tract differences, talker differences can be removed, and the listener parses the signal after removing variation. Similarly, Nearey's log-mean normalization (1978) removes anatomical differences in human speakers but preserves phonemic and sociolinguistic information. The procedure is using information across different vowels and therefore formants. Typically, the frequencies of the first three formants can be used to remove anatomical differences in speakers. The left-over variation is either phonemic or sociolinguistic. Nearey (1989) also suggested that extrinsic factors, like vocal tract size information (corresponding to F3), are considered in conjunction with intrinsic vowel information cues like F0 and formant ratios. Speakers must not only adapt to the changes in acoustic information, which is dependent on the speaker, but also on factors like dialectal variation (Clopper and Tamati 2010, Llopart and Simonet 2017), non-native accents (Bradlow and Bent 2008, Clarke and Garrett 2004), and other idiosyncratic speaker characteristics (Kraljic, Brennan, and Samuel 2008).

While the previous approaches deal specifically with how a human listener could normalize variant inputs, Lobanov (1971) developed a method to improve automatic classification of vowel sounds with the aim to obtain higher correct machine classifications of different natural speech tokens. His method uses a z-score, for this the

mean and standard deviation of a given formant of a given speaker is needed as estimated from several vowel tokens. New information is then categorized by subtracting the formant value in question from the mean formant value and then dividing this result by the standard deviation of the mean (Escudero and Bion 2007).

Disner (1980) compared different normalization methods (Gerstman (1969), Lobanov (1971), Nearey (1978), and Harshman (1970)) using vowel data from different languages (English, Norwegian, Swedish, German, Danish, and Dutch). Her results show that Nearey's method was the most effective method of scatter reduction for all languages with Lobanov and Gerstman being slightly less effective.

Talkers choose different styles of speaking as markers (e.g., of social, gender, dialectal information) and a normalization method that accounts mainly for vocal tract differences cannot account for within-speaker variance. Magnuson and Nusbaum (2007) developed a theory of normalization that uses an active cognitive mechanism to decode the speech signal in real-time. The system monitors and modifies the output depending on the contextual information available (e.g., coarticulation, indexical speaker characteristics, etc.). /I/ might have the same formant pattern for one speaker that /ε/ has in a different speaker. Their model reflects the many-to-many mapping between acoustics and phonetics. To test their theory, they looked at mixed-talker and blocked-talker conditions and found that the advantage of the blocked-talker condition disappeared, when subjects were not aware of the different fundamental frequencies of the voices, showing that talker normalization is an active open-loop process, that is multimodal and multidimensional. Similarly, Johnson (1990) presented listeners with a *hood-hud* continuum with two different F0 contexts. Listeners judged stimuli differently

based on the F0 context, showing that perceived speaker identity is triggering normalization processes.

1.4.2 Vowel normalization as perceptual constancy

A more recent approach suggests that vowel normalization is similar to other areas of perceptual constancy and that social, indexical, and linguistic information all play a role in arriving at a speech signal interpretation. Barreda (2020) suggests that vowel normalization is similar to size constancy in visual perception. Starting from the constant ratio hypothesis, which suggests that listeners can perceive vowels when they vary according to a single proportional parameter for all formants, differences in phonemes can be interpreted according to contextual factors. While objects in the visual domain are scaled according to distance, spectra are scaled according to vocal tract length. Subphonemic variation is not removed as it is used for indexical and sociolinguistic information. Studies (c.f. Charlton et al. 2007, Charlton et al. 2009, Charlton et al. 2012) have shown that associations between size and acoustics might be fundamental in auditory perception in mammals. Different formant patterns are not analyzed as different qualities of sound but attributed to different sizes in speakers, translating to speaker size and vowel quality being orthogonal. Uniform-scaling is phone preserving as the acoustic variability is mapped to change in speaker size and not change in linguistic interpretation.

This theory is an elaboration of Nearey's uniform scaling model, showing that perceived vowel quality is based on the perceived location of the vowel within a speaker's vowel system. Talker characteristics have an indirect effect on the perceived vowel quality and removing all this information as Lobanov suggests, removes more within and between speaker information than human listeners do. While this works well for machine perception, it does not reflect what humans are doing. Barreda's perceptual constancy account suggests that indexical speaker information is not just noise and instead aids listeners in making sense of the non-invariance present in the speech signal. Between-speaker variation is interpreted as variation in the vocal tract, based on uniform scaling. Vocal tract characteristics then inform the interpretation of indexical information about the speaker.

Listeners use context cues to arrive at an interpretation of speech sounds, removing all variance in the signal is unlikely to be what humans are doing. Information like height, gender, and age can help a listener to set up a reference system of sounds for an individual speaker and therefore deal with the variability present in the acoustic signal. It does not make sense to suppose that listeners want to normalize for indexical features, male and female formant patterns differ and this difference in gender helps the listener interpret the acoustic signal accordingly. The same goes for within-speaker differences. Contextual cues can help recalibrate and deal with a different input signal that can nonetheless be mapped onto one interpretation of a phoneme.

1.5 Approach used for this dissertation

This dissertation will adopt a mix of Nearey's pattern recognition approach and probabilistic models, assuming that listeners are sensitive to acoustic patterns associated with vowel identity but can adapt to variant inputs by comparing the similarity of a novel input to the multi-dimensional representations of phonemes in memory and make a decision based on similarity. Therefore, while German and English vowels show different acoustic patterns, English listeners can categorize novel German inputs based on the similarity to the English acoustic patterns stored, showing which acoustic cues in the signal are important to determine similarity.

In the production part of this dissertation, the vowels will be normalized using the Nearey 2-formant extrinsic method. The formula for this is:

$$F_{n[V]}^* = \text{anti-log}(\log(F_{n[V]}) - \text{MEAN}_{\log})$$

Where $F_{n[V]}^*$ is the normalized value for $F_{n[V]}$, formant n of vowel V , and MEAN_{\log} is the log-mean of all F_1 s and F_2 s⁹. This method performs well when the whole vowel system is included, which is the case in this dissertation. The resulting representations of the vowel systems will allow comparisons along the F1/F2 plane and predictions for resulting confusion in the perception of German vowels by English listeners will be made based on the vowel spaces.

Assuming that listeners are sensitive to acoustic patterns and are able to handle variable inputs by comparing their patterns to existing phonemic categories, the first experiment will establish the acoustic patterns of German and American English vowels and their respective language-specific cues used in perception by using statistical

⁹ See http://lingtools.uoregon.edu/norm/norm1_methods.php, last accessed 11/04/2021.

modeling and an identification and rating task of German vowel sounds by naive English listeners to test whether the statistical predictions hold up in human perception of foreign speech sounds.

Chapter 2

In order to look at the role of quantity and quality in German long-short vowels it is important to first establish the acoustic features present in production. The first experiment establishes the acoustic qualities present in the production of German vowels and the saliency of these cues in perception by naive listeners. This is an important first step to establishing the saliency of acoustic cues present in German long short vowels. English is a closely related language to German but has lost the quantity contrasts present in German over time. Additionally, the vowel system is also smaller and shows less overlap spectrally, which is why American English listeners are likely to pick up on spectral differences between German long short vowel pairs. If these vowels differ mainly based on duration, naive listeners might collapse them into the same native English category. However, if spectral differences are salient, listeners are more likely to exploit these differences and perceive the long and short vowels as distinct and therefore map them to different American English vowels.

2.1 Experiment one

The first experiment investigates the production of German vowels by native German speakers and the perception of these vowels by American English-speaking listeners with no prior experience with German. In order to assess the relative importance of spectral form and quantity in the discrimination of vowels in both German and American

English, a production study was conducted in which native speakers of German produced the 16 German monophthongs /i:/ /y:/ /u:/ /ɪ/ /ʏ/ /ʊ/ /e:/ /e/ /ø:/ /o:/ /ɛ/ /ɛ:/ /œ/ /ɔ/ /a/ /a:/ and native speakers of American English (referred to as AE in the following) produced the English monophthongs /ɛ/ /æ/ /e/ /i/ /u/ /ɑ/ /ɪ/ /ɜ:/ /ʊ/ /ʌ/ /o/. Productions were used to establish vowel spaces for both German and American English and the relevance of acoustic features (F1, F2, F3 at midpoint and duration) was assessed using linear discriminant analyses. In a second step, a perception study was carried out in which American English listeners were presented with the German monophthongs in an identification and rating study. Based on these results, vowel-specific Bayesian logistic regression models were run for each German long-short vowel pair that was not perceived as the same American English vowel to determine which acoustic features listeners used in their decision making, to further confirm the importance of quality versus quantity. Specifically, the acoustic similarities and dissimilarities between German and English vowels should predict the perceptual patterns, especially with regard to the different combinations of acoustic features used in vowel disambiguation; because German vowels differ in quantity and quality, it is assumed they can be disambiguated by a different set of acoustic cues than English vowels. American English listeners may rely mainly on F1 and F2 based on their learned L1 contrasts. If this is the case, and German long-short vowel pairs are spectrally different in a salient way, listeners might not collapse the long-short vowel pairs into a single category and instead interpret them as different categories.

Additionally, the data can give some insight into the presence of duration as a cue in the production of German vowels: even though quantity is not an active phonological contrast in American English (c.f. Lindsey 1990, Nishi et al. 2008, McAllister et al. 2002), duration is a salient contrast in the production of long/short German vowel pairs, therefore listeners are predicted to use duration as a secondary cue and less likely merge the long-short pairs into the same AE category perceptually. Duration could act as an enhancing predictor besides formant frequencies (referred to as FFS in the following), even when the competing AE categories are not varying in duration. Therefore, while duration will likely be an important cue for the perception of long-short German vowels, even in non-native listeners. If, however, English listeners will not include duration in their assessment of German vowels, we can expect that the long-short German vowel pairs will be collapsed into one single vowel category in English listeners for German vowel pairs that do not show significant spectral differences.

2.1.1 Background

In the first experiment acoustic features of German vowels and their usage in speech perception are explored. This is one of the first studies systematically exploring the role of both duration and spectral cues in German L1 perception. In line with the speech perception theories presented in section 1.3 in chapter one, statistical models are used to build models of cross-linguistic vowel perception. Since spectral and durational cues

might not be used in the same way by German and English speakers, cue weighting and resulting identification patterns of German vowels should show different outcomes for the L1-English and L1-German speakers, resulting in a confusion matrix based on acoustic cues from production. To understand the results of non-native vowel perception, a good understanding of the L1 vowel space is necessary. Therefore data on the production of American English vowels by monolingual speakers is also collected. The aim is to establish the patterns of non-native and native perception, which can serve as a basis for improved L2 acquisition theories as well as feature engineering for ASR (automatic speech recognition) algorithms.

The data from Experiment 1 provides insight into both the contrastive phonetic properties of English and how much speech perception is constrained by the acoustic patterns of the L1. In addition, information obtained about the degree to which certain acoustic cues are used by listeners could inform both automatic speech recognition and speech synthesis, improving classification accuracy for accents and dialects using patterns from production in conjunction with confusion matrices obtained from human perception and perceptual distance. Furthermore, perceptual data can be used in speech synthesis to improve the quality of the output by including cues that human listeners utilize.

In the experiment, German listeners produced non-words in a /bVt/ structure. From these tokens, the German vowel space was established. Adult L1 English listeners with no previous experience with German were presented with the German tokens in an identification and rating task. Additionally, the same listeners also produced

non-words in order to establish the American English vowel space. This will be described in more detail in the Methods section below.

In trying to understand how non-native speakers perceive foreign sounds, it is important to review the different theories of Second Language Acquisition. In the following, a short review of SLA theories and models as well as an overview of the acoustic properties of California vowels will be given. On this basis, predictions about the German vowel classifications by L1 English speakers will be made.

2.1.2 Second Language Acquisition Theories and models

Second language learning is a very well-researched field in linguistics and over the years many different theories have been explored in an attempt to explain the underlying processes of language learning. The discussion below is an overview of the most influential theories and models, which are relevant here to relate the categorization of unfamiliar speech sounds in perception to the distribution of acoustic properties from speech production in the L1. This section is not intended to be a general review of SLA theories, but a short overview of the origins of SLA theories and those most relevant to this dissertation, specifically the Speech Learning Model (SLM) and the Perceptual Assimilation Model (PAM).

One of the earliest attempts at explaining how speakers learn a language has been made with a theory borrowed from the field of psychology, Behaviorism, which influenced many other disciplines in the 1950s and 1960s. In the tradition of Skinner

and Bloomfield language learning is nothing but another learning process. Within this framework stimuli and responses are the main motors for learning, humans react to given stimuli and experience reinforcement if the reaction is fit for the situation. By constant reinforcement, habits are formed, which is essentially the basis for the learning process of a new skill. If this concept is applied to second language learning, a learner will eventually learn how to produce and perceive L2 sounds if the conversational partner understands and responds. Therefore, producing and perceiving sounds unsuccessfully leads to failed communication or misunderstandings and will not be reinforced. However, successful production and perception of non-native sounds will successfully communicate a speaker's goal and therefore be reinforced and eventually become a habit (c.f. Bloomfield 1933, Skinner 1957, Mitchell et al. 2013, Menezes 2013).

Of particular interest for this dissertation are two models concerned with the acquisition of phonology and phonetics, Flege's (1995) Speech Learning Model (SLM) and Best's (1995) Perceptual Assimilation Model (PAM). Both models address the question of how learners in different stages categorize novel L2 sounds based on their existing phonological system in the L1. Flege's (1995) speech learning model (SLM) proposes that it is easier to perceive L2 sounds that are dissimilar from any existing L1 categories. In the SLM sounds can be identical, similar or new (as front rounded vowels would be for L1 English speakers). A new category only emerges for those tokens that are most dissimilar and do not fit into the L1 system. For both the identical or similar tokens learners would not establish a new category, which could make the perception and production of these sounds in an L2 harder. However, research has shown that the

category transfer from the L1 is not direct, instead sounds that have correlates in the L1 may be even harder to produce in a nativelike manner than those sounds that do not have any correlates and require a new category (Bohn and Flege 1992, Flege 1992, 1995). The influence of the first language on the second language can manifest itself as an accent because of subtle phonetic differences in the production of L2 vowels (O'Brian and Smith 20210). Some experimental evidence comparing English and German comes from O'Brian and Smith (2010), who show that while English learners of German produced German /u:/ with higher F2 values than they did the North American English /u/, they retained formant patterns that were close to their native variety of English. It is notable that German /u:/ has lower F2 values than the English /u/, so producing it with higher F2 values might be a case of overcorrection or hyperproduction. An important difference between the two vowel systems is that German has front rounded vowels /y:/, /ʏ/, /ø:/, and /œ/, which English lacks. In their study all subjects produced German /y:/ differently from /u:/, indicating a separate category for this vowel despite its lack in their native vowel systems. Polka (1995) looked at English listeners' perception of the German front rounded vowel contrasts /y/ and /u/ and /y:/ and /u:/ in a discrimination task and found that the English listeners mapped those four vowels to high back rounded L1 vowel categories. She also found that subjects had more trouble categorizing the short vowels since English does not have a phonological long-short distinction. Similarly, Bohn and Flege (2004) have shown that German learners' perception of English /æ/ only improved with prolonged exposure to the target language and merged with the closest native category /ɛ/, whereas the perception of the contrast

/i/ and /I/ was not affected by this learning curve, as this contrast exists in German as well.

The PAM is based on the direct realist approach of speech perception mentioned earlier and assumes that listeners directly perceive articulatory gestures from the speech signal. It describes naive listeners' perceptions of foreign phonemes based on five different categories: two category, category goodness, single category, uncategorizable-categorizable, both uncategorizable. In the two-category case, a new L2 sound assimilates to an existing L1 category that is different. In the category goodness case two L2 sounds assimilate to one L1 sound category, for example, both German /u:/ and /ʊ/ assimilate to English /u/ with one being accepted as a better fit whereas the other is just an acceptable fit. In the single category case, the same thing happens, except now both L2 sounds are not a good fit. In the uncategorizable-categorizable case, only one L2 sound assimilates to an L1 category and the other one does not. In the both uncategorizable case, no L2 sound will merge into any existing L1 category. Similar to the SLM, the PAM also concludes that sounds will be easier to perceive when they are least similar to existing L1 categories.

As both models deal with the question of similarity in speech sounds, it is important to first establish an acoustic space for perceptual similarity of German and English based on which possible confusion patterns can be predicted. Strange et al. (2004) looked at the spectral similarities between German and English and predicted listener behavior based on their findings. The German long vowels posed a problem for English listeners in perception and were instead mapped to English diphthongs. In the same study listeners also had to give goodness ratings for each vowel, and the German

front rounded vowels received poor ratings overall. In a follow-up study, Strange et al. (2005) presented English listeners with only front rounded German vowels and found that they were consistently mapped onto English back vowels with a medium goodness of fit rating.

Perception and production are highly correlated. The patterns of production in English vowels can be used to predict the assimilation patterns of perception for the L2 German vowels. An illustration of the naive state of cross-linguistic vowel perception is given in Figure 4, in which an L1-English listener categorizes each German vowel based on their L1-English system.

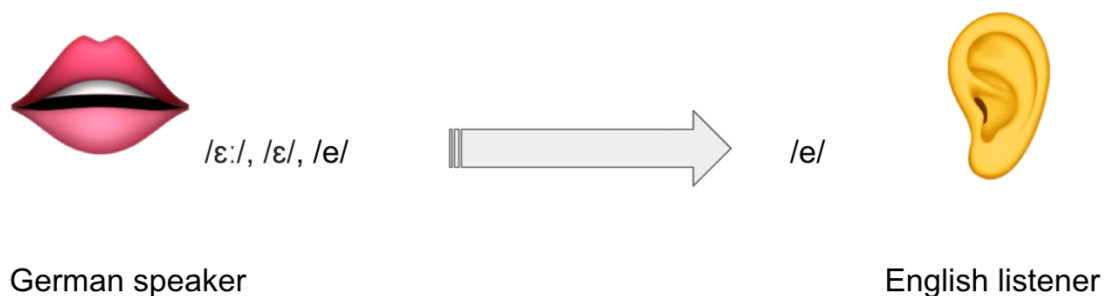


Fig. 4: Possible perceptual patterns for German */ɛ:/, /ɛ/, and /e:/* in native American English listeners

While the three German vowels differ in quantity and quality in the German phonological system, they might be perceived as the same vowel to a naive AE listener based on the phonological system and contrasts of their L1.

2.1.3 California English vowels

Even though there is substantial variation in American English vowels just like with German vowels, the literature largely agrees that the vowel system consists of the monophthongs /i, e, u, o, ɔ, ɑ, ɪ, ε, ʊ, ʌ, æ, a/ and the diphthongs /aɪ, aʊ, ɔɪ/.¹⁰

Tense (Long)			Lax (Short)			Diphthongs		
i	meat	[mit]	ɪ	mitt	[mɪt]	aɪ	ice	[ajs]
e	mate	[met]	ε	met	[mεt]	aw	louse	[laws]
u	food	[fud]	ʊ	good	[gʊd]	oj	voice	[vojs]
o	moat	[mot]	ʌ	mutt	[mʌt]	ju	puce	[pjus]
ɔ	caught	[kɔt]	æ	cat	[kæt]			
ɑ	balm	[bɑm]	a	bomb	[bɑm]			

Fig. 5: American English vowels (Putnam and Page 2020:111)

In terms of quantity, the vowel system is simply described as tense vowels always being long and lax vowels always being short and there is much disagreement about vowel quantity existing in American English. Examples for this distinction are given in Figure 5 above. While Chomsky and Halle (1968) treat the contrast as tense/lax spectrally, with lax vowels being produced more central and tense vowels more peripheral in the F1/F2 plane, Halle (1977) later treated it as a long/short distinction instead. In support of quantity as duration, House (1961) collected speech data of three male speakers, whose vowel productions seemed to be contrastive long-short pairs for a subset of

¹⁰ The German vowel system has been discussed in depth in chapter one. Please refer to chapter one for an in depth overview of the MSG vowel system and its contrasts.

vowels (/i/-/ɪ/, /u/-/ʊ/, /e/-/ɛ/, /ɑ/-/ʌ/). In contrast, Lehiste (1970) classified American English as having only secondary quantity, meaning that spectral information is used as the primary cue.

While previous research has described the vowel system to be largely homogenous in the Western United States based on the bot-bought merger (Labov 1991), recent research has shown California English to show distinct features. The California vowel shift has been shown in both Northern and Southern California. Hinton et al. (1987) documented the fronting of /u/ and /oʊ/. An example of the California vowel space is given in Figure 6.

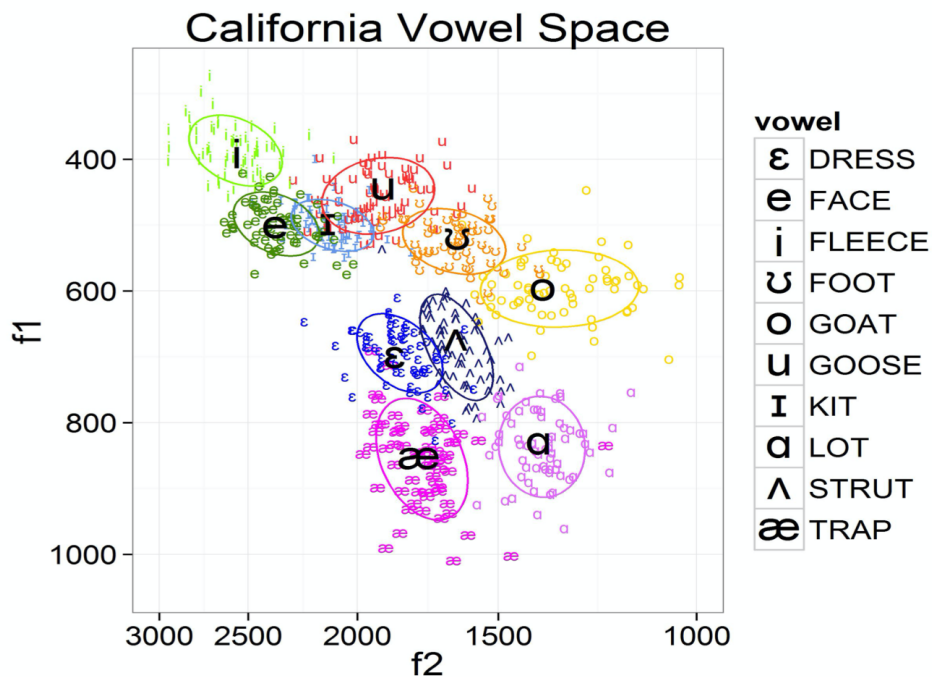


Fig 6: California vowel space (Holland 2014:58)

While Holland (2014) also found /o/ and /ɪ/ to be fronting, he additionally showed that /u/ and /ʌ/ are moving upward in the vowel space. He judges the /u/ fronting as stable, and the /o/ and /ɪ/ fronting as still in progress. Another shift Holland shows is rounding, with /u/ being rounded but /ɪ/ not showing rounding, which could make rounding a distinctive feature between the two vowels. In addition to fronting, Hickey (2018) also found lowering of the short front vowels causing the vowel in DRESS to be realized as /æ/, the vowel in KIT as /e/, and the vowel in TRAP as /a/. He mentions the exception of /æ/ in pre-nasal context, where the vowel is raised instead of lowered.

In comparison to the German vowel space, /u/, /ʊ/, and /o/ inhabit the space that the German front rounded vowels (/y:/, /Y/, /ø:/, /œ/) take up, so L1 English listeners might categorize these sounds based on the distribution of native sounds inhabiting this area in the vowel space.

2.2 Materials and Methods

2.2.1 Participants

2.2.1.1 German

21 native speakers of German were recruited via Facebook for this study (m = 10, f = 11, mean age = 49.9, age σ = 10.9) and completed the experiment on Qualtrics. All participants reported German as their native language and all but three reported not having any experience with any other language. None reported any problems with the

experiment platform. All participants except 3 reside in North Rhine-Westphalia. A table showing detailed demographic information can be found in the appendix.

2.2.1.2 English

All participants were L1 English speakers with no experience in German. All participants were undergraduate students of the University of California and were recruited from the psychology undergraduate research pool through SONA. Participants completed the entire experiment from their own homes on Qualtrics. Those who did not report English as their dominant language or did not complete the recording part correctly were excluded. After this, a total of 11 speakers were left (f = 6, m = 4, non-binary = 1, mean age = 26, age σ = 8.94). All but two speakers had experience with other languages than English. All participants reside in California. A table showing detailed demographic information can be found in the appendix.

2.2.2 Production stimuli

2.2.2.1 German

Participants recorded 16 target items. All items had either a /bVtt/ structure, to indicate a short vowel, or a /bVht/ structure, to indicate a long vowel. The target words included the closed vowels /i:/ /y:/ /u:/ /ɪ/ /ʏ/ /ʊ/, the mid vowels /e:/ /e/ /ø:/ /o:/ /ɛ/ /ɛ:/ /œ/ /ɔ/ and the open vowels /a/ /a:/. The reason for including /ɛ:/ in this study, even though some

researchers like Moulton (1962), Sanders (1972), and Reis (1974) have not accepted its existence, is that Frank (2021) has shown /ɛ/ and /ɛ:/ to be only partially merged in the region the speakers for this experiment were recruited from. Furthermore, Predeck et al. (2021) have shown productions for the two vowels to be significantly different in F1 for speakers from the same region. Thus, it appears that at least some varieties of German have a clear and stable contrast between the two. Therefore in the following, the added distinctions of /e:/-/e/ and /ɛ:/-/ɛ:/ in the German vowel system will be used.

2.2.2.2 English

Participants recorded 11 target items. All items had a /bVd/ structure. The target words included the closed vowels /ɪ/, /i/, /ʊ/, and /u/, the mid vowel /ɛ/, /e/, and /o/ and the open vowels /æ/, /ʌ/, /ɑ/, and the r-colored vowel /ɜ:/.

2.2.3 Procedure

2.2.3.1 Production

2.2.3.1.1 German

The experiment was conducted fully online, using the Qualtrics survey platform. Participants were instructed to sit in a quiet room and use a computer to complete the experiment. Before starting their recordings, participants were asked to record a test

sentence to ensure that their microphone worked and sound was picked up and saved to the AWS cloud. After the test recording, participants were instructed to read a list of 31 sentences, containing the 16 target items as well as distractor items. The target items were presented in a carrier phrase: “ ____, *ich sage das Wort ____ zu dir.*” (“ ____, *I say the word ____ to you.*”), to ensure natural prosody. After finishing the recording portion of the experiment, participants were asked to fill out a background questionnaire.

The sentences were written in Modern Standard German using German spelling conventions to indicate long or short vowels. A short vowel preceded a double consonant and a long vowel preceded an /hC/ cluster, so target words were presented in /bVht/ and /bVtt/ frames. The following example shows the first four sentences, including two target items and two distractor items.

1. Tos, ich sage das Wort tos zu dir.
2. Bött, ich sage das Wort bött zu dir.
3. Buht, ich sage das Wort buht zu dir.
4. Kieb, ich sage das Wort kieb zu dir.

2.2.3.1.2 English

The full experiment was hosted on the Qualtrics online survey platform. Prior to the production part of Experiment 1, participants were given instructions and then they completed a practice recording with two sentences to ensure that their microphones worked. After completing this, participants were asked to produce 11 non-words with a /bVt/ structure in a carrier phrase (“ ____, the word is ____”). For each non-word, a real English rhyme word was presented to the speakers as well to ensure that the correct

target vowel would be produced (“*target word*” rhymes with “*rhyme word*”). Participants were instructed to not read the sentences in parentheses out loud as this only served as a guide to the target vowel production. Participants only read the list once, producing two tokens of the target word. These tokens were used to establish the features of the English vowel space in order to compare these to the German tokens. The prompts looked like this:

1. (Food rhymes with bood.) Bood, the word is bood.
2. (Bud rhymes with gud.) Gud, the word is gud.
3. (Sad rhymes with gad.) Gad, the word is gad.
4. (Bird rhymes with gird.) Gird, the word is gird.

2.2.3.2 Perception

After completing the recording procedure, participants were presented with two perception tasks, a forced-choice word selection and a goodness of fit rating task. In total, participants completed 144 trials (16 vowels, 9 German speakers, $m = 4$, $f = 5$). In a given trial, participants heard one of the German stimulus words. First, they completed the categorization where they were presented with word options in a /bVt/ structure, as all German words were recorded in this structure, on the screen and prompted to click the one that they thought to be most similar to the stimulus.

Second, they were asked to determine the goodness of fit on a slider bar underneath the word option buttons, with the slider on a scale, starting from 'best' to 'worst' in 7 steps. Hearing the stimulus more than once was not an option. All individual vowel trials were repeated only once per participant. The experiment took an average of 30 minutes to complete and stimuli were presented in randomized order. Figure 7

shows the word choice and rating task participants completed in the perception part of the experiment.

Trial of

What did this word sound like most?

Bood (rhymes with food)

Bud (rhymes with mud)

Bad (rhymes with sad)

Bird (rhymes with third)

Bod (rhymes with cod)

Bed (rhymes with red)

Bade (rhymes with fade)

Beed (rhymes with seed)

Bid (rhymes with kid)

Bould (rhymes with good)

Bode (rhymes with code)

How much like the word you chose did the word sound? From 1 (not at all like it) to 7 (perfectly like it).



Fig 7: Example of Perception Experiment Template

Listeners saw the same template as shown in Figure 7 for all trials and had to make a word and rating selection before they could move on to the next trial. After completing the perception task participants filled out a demographic questionnaire.

2.2.4 Acoustic measures

Target words were forced aligned using an online version of MAUS¹¹ to generate Praat TextGrids. The TextGrids were hand-corrected for the start and the end portions of the target vowels.

An adapted Praat script (Lennes 2003) was used to measure F1, F2, F3, f0, intensity at 20%, 50%, and 80%, and duration of the vowels. Formant measures were normalized using the `normalize` function with the `neareyE` parameter in `phonTools` (Barreda 2015). Vowels were plotted using the *phonR* package (McCloy 2016) and normalized and plotted using NORM v.1.1.¹²

2.2.5 Statistical Analysis

To investigate the relevance of the acoustic features measured, two linear discriminant analysis (LDA) models were run on a subset of English vowels (/æ/, /i/, and /ɑ/) and a

¹¹ <https://clarin.phonetik.uni-muenchen.de/BASWebServices/interface/WebMAUSBasic>

¹² <http://lingtools.uoregon.edu/norm/norm.php>

subset of German vowels (/a:/, /i:/, and /o:/).¹³ Predictors included F1, F2, and F3 at onset, midpoint, offset (normalized using the neareyE method in phonTools), and duration. Because features were on very different scales, all features were scaled using the *sklearn StandardScaler* (Pedregosa et al. 2011), which normalizes each feature column, resulting in each column having $\mu = 0$ and $\sigma = 1$. Both models were cross-validated using the *sklearn cross_val_score* (Pedregosa et al. 2011) method and an L2 penalty term was added to the model to avoid overfitting.

To analyze the importance of duration and spectral information in the perception of non-native vowels, vowel-specific models were run for each long/short German vowel pair that was not perceived as the same AE vowel in R using the *brms* package (Bürkner, 2016). German spectral measures at the midpoint (normalized) and duration were included as predictors. Responses were coded as binary (which AE was perceived). The data were analyzed using Bayesian binary logistic regression models in R with the *brms* package. Main effects included spectral measures and duration. Random effects included by-German Speaker random intercepts and by-AE Listener random intercepts. The priors used were a student's t-distribution ($\nu = 3$, $\mu = 0$, $\sigma = 3$) for the regression coefficients, and a student's t-distribution ($\nu = 3$, $\mu = 0$, $\sigma = 2.5$) for standard deviations of random effects. All models converged (Rhat = 1.0). (*brm* syntax: F1_50 + F2_50 + F3_50 + duration + (1|ID_GER) + (1|ID_AE), family = bernoulli(logit))

¹³ These subsets were used because both datasets contained a high number of categories but relatively little data points so running the classification models on all classes would lead to low accuracies.

2.2.6 Results

2.2.6.1 Production

2.2.6.1.1 German

To investigate the relevance of the acoustic features measured, a linear discriminant analysis (LDA) model was run on a subset of German vowels (/a:/, /i:/, and /o:/). Predictors included F1, F2, and F3 at onset, midpoint, offset (normalized), and duration. Mean prediction accuracy of the LDA over 10 folds reached 92%.

Figure 8 shows the resulting normalized vowels for F1 and F2 in Hertz as taken from vowel midpoints at 50%. The figure shows that most long and short vowel pairs are spectrally different and inhabit different areas of the vowel space. This has been verified by separate linear mixed-effects models run on F1 and F2 for each vowel pair.¹⁴ Exceptions to this are /u:/-/ʊ/ and /ɛ:/-/ɛ/, which are spectrally closer together in the F1/F2 plane. In the case of /ɛ:/-/ɛ/, only F2 was significantly different ($p < 0.001$). This could be due to the fact that /ɛ:/ is undergoing a merger with /e:/ and speakers from the area observed in this dissertation are in an area where this merger is still incomplete (Frank 2021). In the case of /u:/-/ʊ/, only F1 was significantly different ($p < 0.001$). The minimal differences seen in this vowel pair are likely due to a flaw in the experimental design. The nonwords shown to participants for this pair were *buht* and *but*. The lack of a double consonant at the end of *but* (instead of *butt*) could have caused German speakers to interpret the vowel as somewhere between long and short as spelling convention indicates a short vowel when a double consonant follows.

¹⁴ The model outputs are fully reported in the appendix.

It is noteworthy that /a:/-/a/ are spectrally different in both F1 ($p < 0.001$) and F2 ($P < 0.001$): this is in contrast to most of the literature reviewed earlier in this dissertation that states that /a:/-/a/ differ durationally but not spectrally.

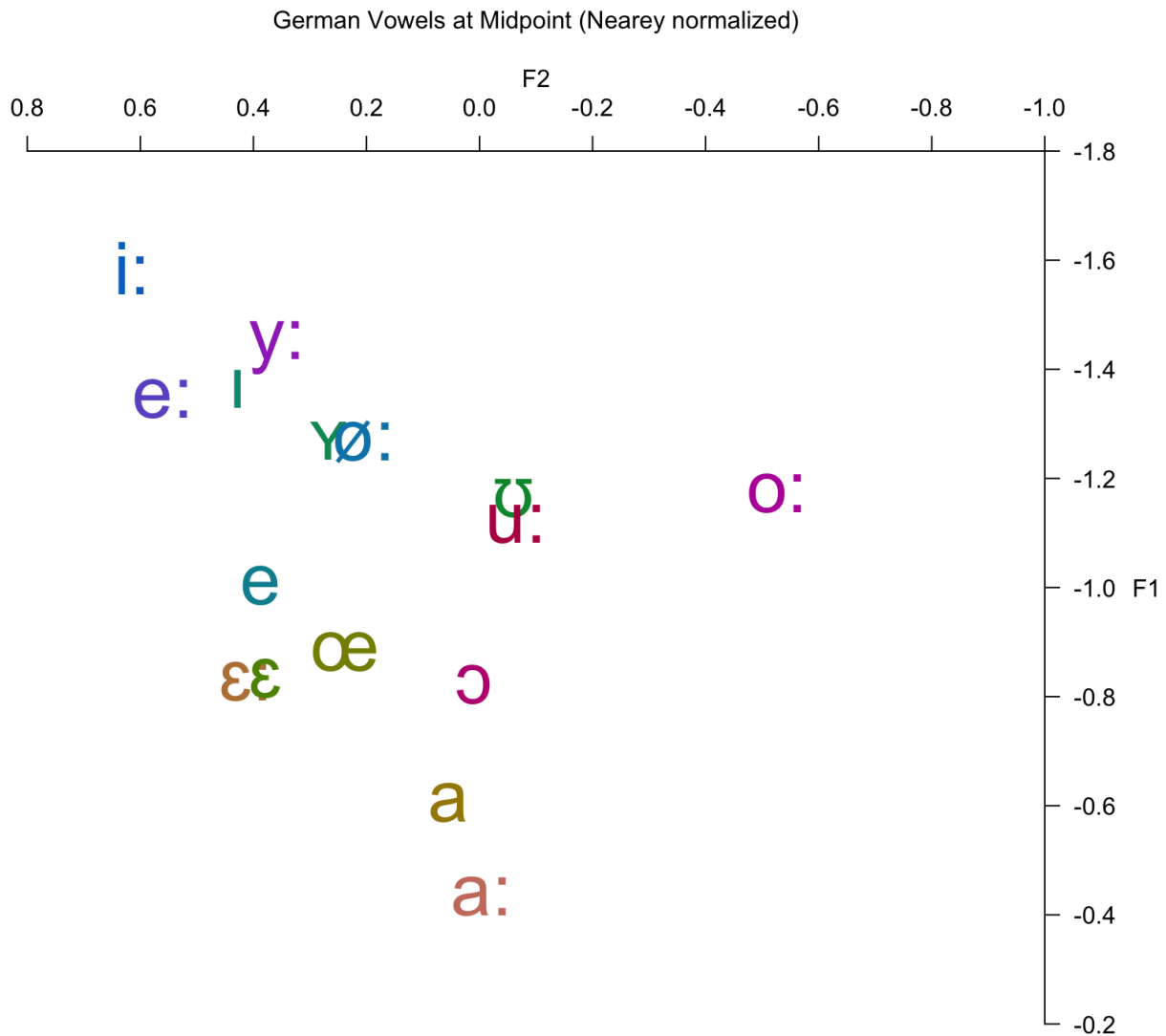


Fig. 8: Normalized German vowels for 21 speakers from experiment one

In addition, all vowels show differences in duration with the short vowels being between 35 and 25 percent shorter than the long vowels as depicted in Figure 9. The exceptions

are /ʏ/ and /y:/, in which the shorter vowel is only 15 percent shorter than the long vowel, and /u:/ and /ʊ/, in which the shorter vowel is only 8% shorter.¹⁵ A linear mixed-effects model¹⁶ confirmed that there is a significant difference in duration between long and short vowels ($p < 0.001$).

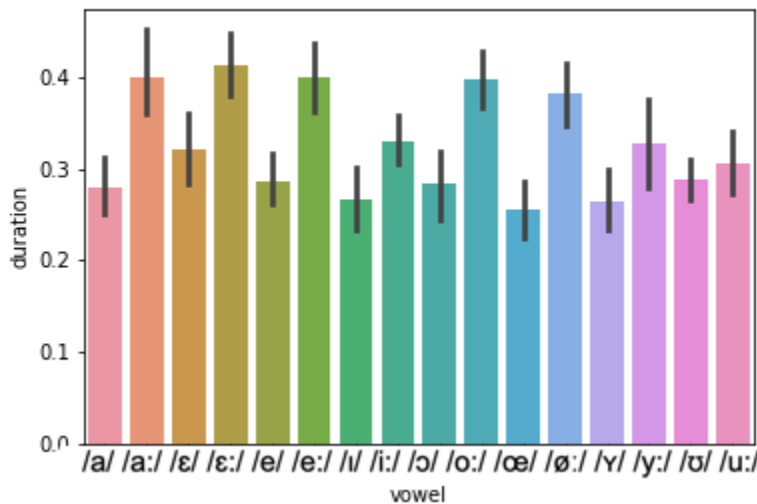


Fig. 9: Vowel duration (in seconds) for German

Additionally there is a pattern for German vowels to not simply show overall longer durations for the long vowels, but rather vowel specific duration patterns as shown in Figure 10. This indicates that long vowels are not simply about the same length but that there are different individual durations of long vowels.

¹⁵ As mentioned above, the minimal differences seen in /u:/ and /ʊ/ are attributed to a flaw in the experimental design where the nonwords shown to participants for this pair were buht and but. The lack of a double consonant at the end of but (instead of butt) could have caused German speakers to interpret the vowel as somewhere between long and short as spelling convention indicates a short vowel when a double consonant follows.

¹⁶ Lmer syntax: $\text{duration} \sim \text{Length} + (1 | \text{ID_GER})$ where Length was coded as binary (0 =short, 1 = long)

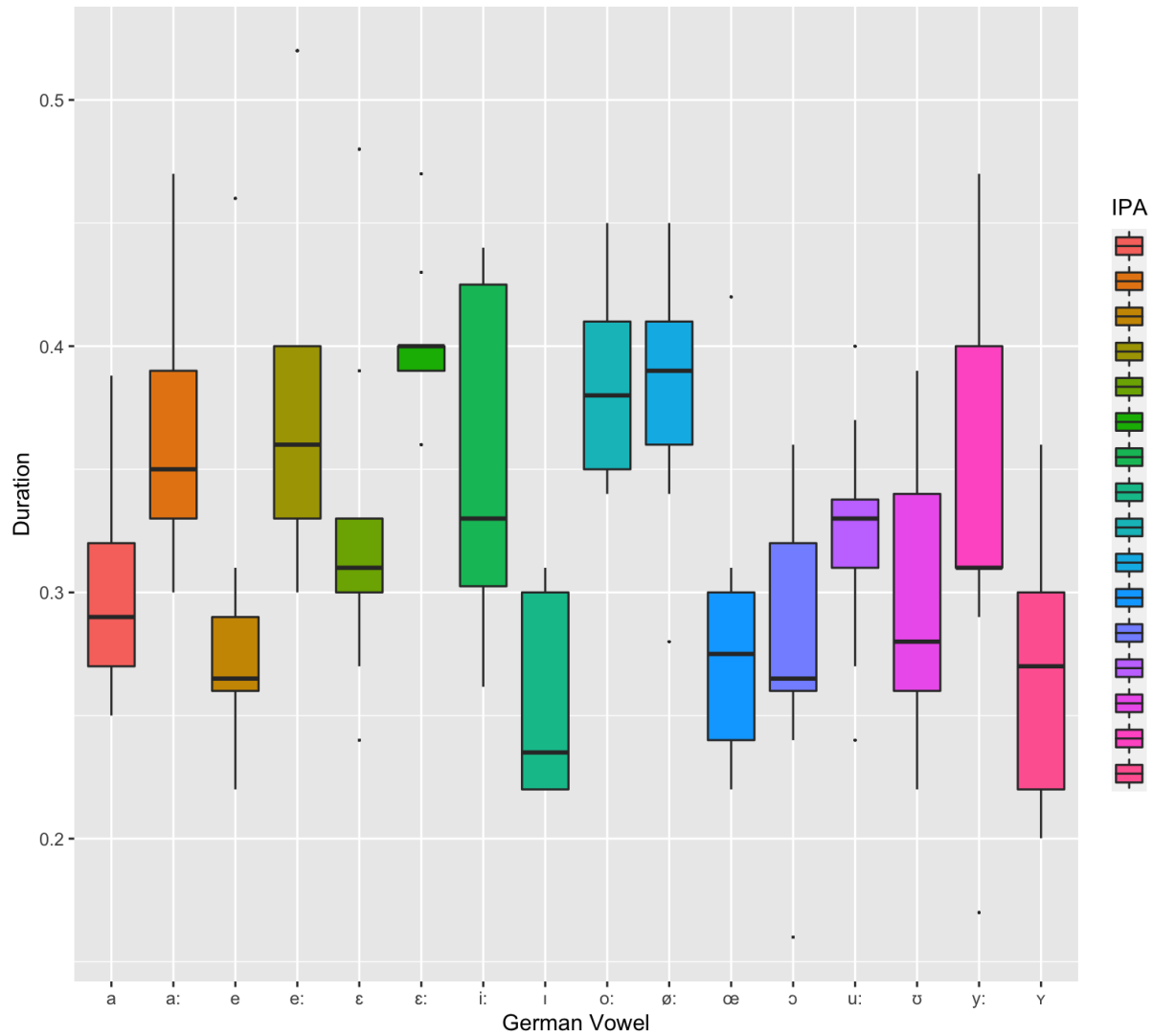


Fig. 10: Quantity (in seconds) of German long-short vowels

2.2.6.1.2 English

To investigate the relevance of the acoustic features measured, a linear discriminant analysis (LDA) model was run on a subset of English vowels (/æ/, /i/, and /ɑ/).

Predictors included F1, F2, and F3 at onset, midpoint, offset (normalized), and duration. Mean prediction accuracy of the LDA over 10 folds reached 95%.

Figure 11 shows the resulting normalized vowels for F1 and F2 in Hertz as taken from vowel midpoints at 50%. The figure shows that all vowels are spectrally different and inhabit different areas of the vowel space.

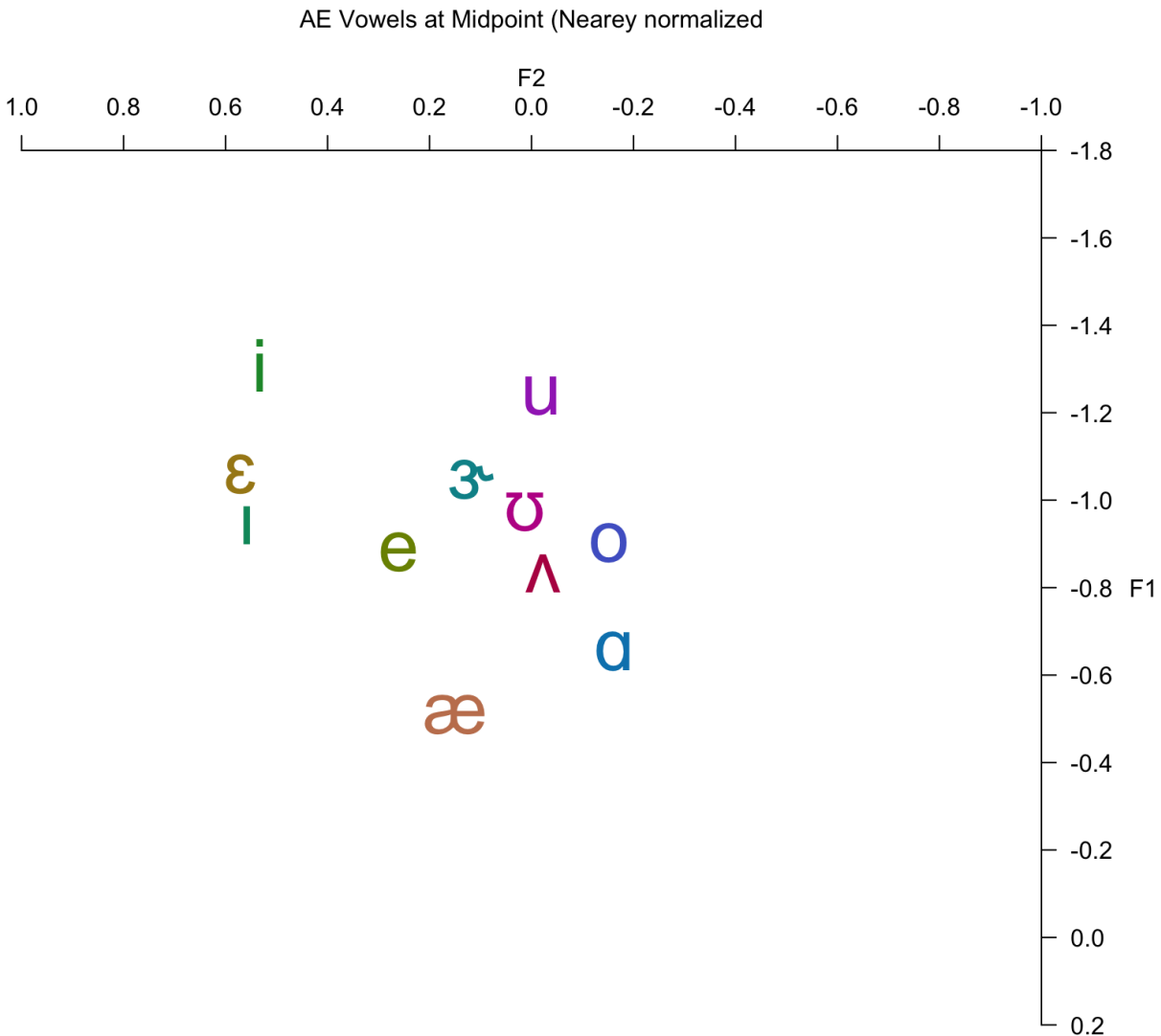


Fig. 11: Normalized English vowels for 11 speakers from experiment one

All English vowels are on average longer in duration (between 411 ms and 570ms) than the German vowels (between 263ms and 410 ms), especially the German short vowels (between 263ms and 314ms). This could influence the perception of the German long-short vowel pairs by English listeners in such a way that duration might be ignored completely, or only used as a cue when spectral cues alone are not sufficient. Average durations for English vowels are shown in Figure 12 below.

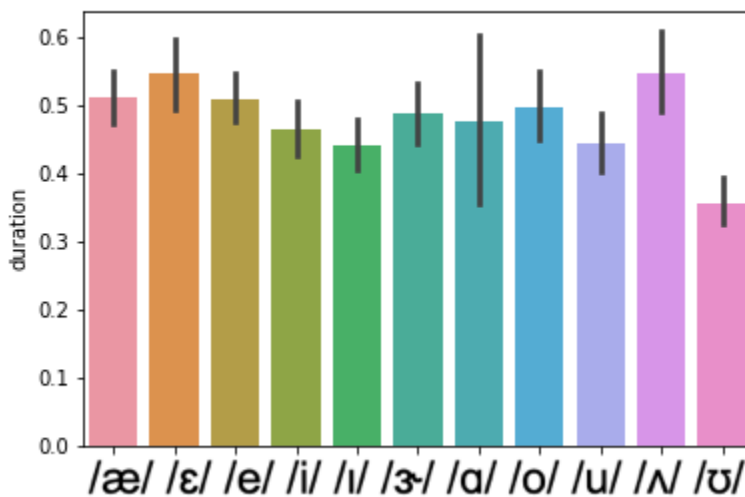


Fig. 12: Vowel durations (in seconds) for English

2.2.6.2 Perception

2.2.6.2.1 Human Listener Perception

The results show that in the perception task most German vowels were mapped to two potential English vowels and most were not rated as perfect fits. Table 6 shows only the

highest percentage for each American English classification, additionally average rating and closest AE based on Euclidean distance is given.

Table 6: Classifications of German vowels as English vowels, rows in gray show that AE listeners selected the closest AE vowel

German Word Context and Vowel	English Word and Perceived Vowel (% of classification)	Average Rating (1 worst, 7 best)	Nearest AE vowel based on Euclidean distance (F1, F2, F3)
Baecht /ɛ:/	Bed /e/ (61.6%)	7	/e/
Baett /ɛ/	Bed /e/ (62.6%)	5	/e/
Baht /a:/	Bad /æ/ (59.6%)	7	/ɑ/
Batt /a/	Bud /ʌ/ (39.4%) Bad /æ/ (36.4%)	5 5	/ʊ/
Beht /e:/	Beed /i/ (53.5%)	7	/ɛ/
Bett /e/	Bed /e/ (69.7%)	5	/e/
Bieht /i:/	Beed /i/ (82.7%)	5	/i/
Bitt /ɪ/	Bid /ɪ/ (55.4%)	5	/ɛ/
Boecht /ø:/	Bood /u/ (43.6%)	4	/ʊ/
Boett /œ/	Bud /ʌ/ (57.3%)	5	/ʊ/
Boht /o:/	Bode /o/ (36.4%)	6	/ɑ/
Bott /ɔ/	Bod /ɑ/ (48.2%)	5	/ʌ/
Bueht /y:/	Bood /u/ (67.3%)	5	/e/
Buett /ʏ/	Bould /ʊ/ (38.2%)	5	/ʊ/
Buht /u:/	Bood /u/ (61.8%)	4	/ɑ/
But /ʊ/	Bood /u/ (45.5%)	6	/ɑ/

While /u:/ and /ʊ/ and /ɛ:/ and /ɛ/ have been collapsed into the same English categories, most other long-short vowel pairs were classified as different English vowels. If German long/short vowel pairs differ mainly in duration and not spectral patterns, these results can not be explained since all AE vowels were on average longer than the German vowels and there are no comparable length patterns. Instead, the confusion patterns show that each long/short German vowel pair differs not only durationally but also differs enough spectrally to be perceived as different AE vowels, instead of a long/short pair being collapsed into the same category. Table 7 gives an overview of the different confusion patterns for all vowels.

Table 7: Vowel confusion matrix from human perception, rows in blue were perceived as the same AE vowel, cells in green show the majorly perceived AE vowel

	/ɛ/ Bade	/æ/ Bad	/e/ Bed	/i/ Beed	/u/ Bood	/ɑ/ Bod	/ɪ/ Bid	/ɜ/ Bird	/ʊ/ Bould	/ʌ/ Bud	/o/ Bode	Total
/ɛ:/ Baeht	6	25	61	1	0	0	1	3	1	1	0	99
/ɛ/ Baett	2	19	62	3	3	1	2	4	1	2	0	99
/a:/ Baht	0	59	1	0	4	22	0	0	0	12	1	99
/a/ Batt	0	36	10	1	2	9	0	0	1	39	1	99
/e:/ Beht	6	2	12	53	3	0	21	2	0	0	0	99
/e/ Bett	0	2	69	2	1	0	11	8	0	5	1	99
/i:/ Bieht	1	0	1	91	1	0	13	2	0	1	0	99
/ɪ/ Bitt	0	0	9	7	4	2	61	7	10	9	1	99
/ø:/ Boeht	1	0	2	0	48	0	0	11	37	9	2	99
/œ/ Boett	0	4	11	0	6	3	2	7	13	63	1	99
/o:/ Boht	0	0	0	1	33	8	0	1	23	4	40	99
/ɔ/ Bott	0	1	0	0	2	53	0	2	2	44	6	99
/y:/ Bueht	1	0	1	3	74	0	1	4	20	3	3	99
/ʏ/ Buett	1	0	0	0	28	5	3	4	42	27	0	99
/u:/ Buht	0	0	0	1	68	4	0	0	26	3	8	99
/ʊ/ But	1	0	0	0	45	6	0	3	31	11	8	99

2.2.6.2.2 Vowel Specific Models

To further investigate the importance of each predictor in the perception of non-native vowels, vowel specific analyses were performed for each German long/short pair that was not perceived as the same AE vowel. For each vowel pair, Bayesian binary logistic regression models were run using the German acoustic measurements. The dependent variable was the English vowel response given by human listeners for each German vowel within the pair, for example AE /ɪ/ or /i/ for the German /i:/-/ɪ/ pair. For each model posterior predictive checks were performed to check model fit visually using the *pp_check* function from the *bayesplot* R package (Gabry and Mahr 2022). All models were good fits for the observed data. Vowel specific models are reported in the following.

For /a:/-/a/, the effect of F1 at midpoint was -5.61 (95% credible interval [-7.23, -4.12]), the effect of F2 at midpoint was -1.86 (95% credible interval [-3.75, -0.11]), the effect of duration was -6.77 (95% credible interval [-12.28, -1.58]). Table 8 shows all effects.

Table 8: Brm Output for German /a:/-/a/, perceived as AE /æ/ or /ʌ/ (regression reference group AE /æ/)

	Estimate	Est.Error	l-95% CI	u-95% CI
Intercept	-0.77	1.64	-4.54	1.80
F1_50	-5.61	0.79	-7.23	-4.12
F2_50	-1.86	0.93	-3.75	-0.11
F3_50	0.30	1.28	-2.22	2.81
duration	-6.77	2.72	-12.28	-1.58

For /e:/-/e/, the effect of F1 at midpoint was -3.49 (95% credible interval [-4.48, -2.58]), the effect of F2 at midpoint was 12.67 (95% credible interval [9.41, 16.27]), the effect of F3 at midpoint was 7.42 (95% credible interval [2.84, 12.52]). Table 9 shows all effects.

Table 9: Brm Output for German /e:/-/e/, perceived as AE /e/ or /i/ (regression reference group AE /e/)

	Estimate	Est.Error	l-95% CI	u-95% CI
Intercept	-15.46	2.31	-20.33	-11.31
F1_50	-3.49	0.49	-4.48	-2.58
F2_50	12.67	1.77	9.41	16.27
F3_50	7.42	2.50	2.84	12.52
duration	-2.13	1.98	-6.18	1.54

For /i:/-/i/, the effect of F2 at midpoint was -5.72 (95% credible interval [-8.98, -2.82]), the effect of duration was -6.69 (95% credible interval [-11.76, -2.14]). Table 10 shows all effects.

Table 10: Brm Output for German /i:/-/i/, perceived as AE /i/ or /ɪ/ (regression reference group AE /i/)

	Estimate	Est.Error	l-95% CI	u-95% CI
Intercept	7.45	1.62	4.49	10.69
F1_50	0.35	0.44	-0.54	1.18
F2_50	-5.72	1.55	-8.98	-2.82
F3_50	-2.64	1.63	-5.99	0.35
duration	-6.69	2.46	-11.76	-2.14

For /o:/-/ɔ/, the effect of F1 at midpoint was -3.72 (95% credible interval [-5.33, -2.24]), the effect of duration was 8.51 (95% credible interval 1.10, 16.94]). Table 11 shows all effects.

Table 11: Brm Output for German /o:/-/ɔ/, perceived as AE /a/ or /o/ (regression reference group AE /a/)

	Estimate	Est.Error	l-95% CI	u-95% CI
Intercept	-8.51	1.86	-12.40	-5.15
F1_50	-3.72	0.79	-5.33	-2.24
F2_50	0.37	0.67	-0.92	1.68
F3_50	2.28	1.28	-0.13	4.96
duration	8.51	4.12	1.10	16.94

For /ø:/-/œ/, the effect of F1 at midpoint was 3.65 (95% credible interval [2.87, 4.47]), the effect of F3 at midpoint was -6.22 (95% credible interval [-8.04, -4.58]), the effect of duration was -8.52 (95% credible interval -12.22, -5.01]). Table 12 shows all effects.

Table 12: Brm Output for German /ø:/-/œ/, perceived as AE /u/ or /ʌ/ (regression reference group AE /u/)

	Estimate	Est.Error	l-95% CI	u-95% CI
Intercept	10.65	1.12	8.56	12.88
F1_50	3.65	0.41	2.87	4.47
F2_50	0.33	0.37	-0.39	1.09
F3_50	-6.22	0.90	-8.04	-4.58
duration	-8.52	1.84	-12.22	-5.01

For /y:/-/ʏ/, FFS at midpoint did not show an effect. Instead, listeners seemed to

instead use information from the start point of the vowel: the effect of F1 at onset was 1.07 (95% credible interval [0.38, 1.77]). Table 13 shows all effects.

Table 13: *Brm Output for German /y:/-/ʏ/, perceived as AE /u/ or /ʊ/ (regression reference group AE /u/)*

	Estimate	Est.Error	l-95% CI	u-95% CI
Intercept	1.60	0.87	-0.06	3.33
F1_20	1.07	0.35	0.38	1.77
F2_20	0.33	0.28	-0.23	0.87
F3_20	-0.46	0.66	-1.73	0.83
duration	-1.05	1.35	-3.70	1.61

These results show that while naive American English listeners largely rely on spectral features when identifying non-native vowel sounds, they also used duration in half of the vowel pairs.

Additionally, vowel height seems to act as an important factor for listeners with almost all German vowels being mapped onto an English vowel with the same height. Figure 13 shows a comparison between German vowels' and American English vowels' height and frontness. The way that the naive AE listeners used spectral cues is in line with the differences observed in production. For example, German /a:/-/a/ differ a lot in F1, the two AE vowel categories that listeners perceived the German vowel pair as, /æ/ and /ʌ/, also show large differences in F1. In the model, F1 at the midpoint had an effect of -5.61 (95% credible interval [-7.23, -4.12]). So naive listeners are exploiting the cues present in German which are also used in their L1 system to distinguish between two vowels. F1 has been shown to have a larger effect on vowel discrimination (Di Benedetto 1989) and German has more height contrasts, than it has backness

contrasts, as shown in figure 13.

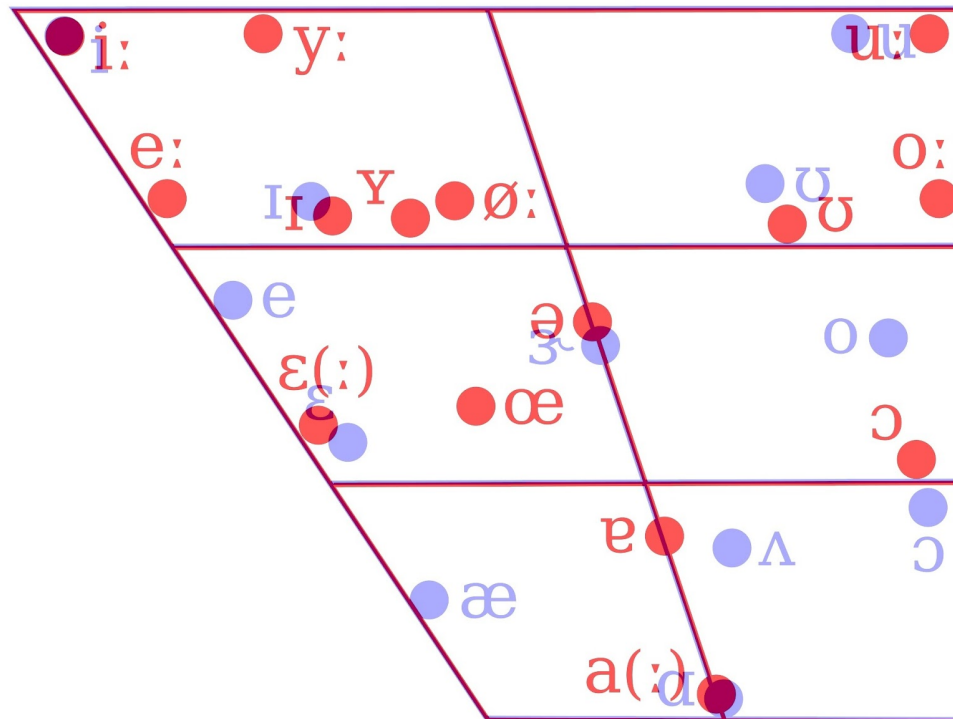
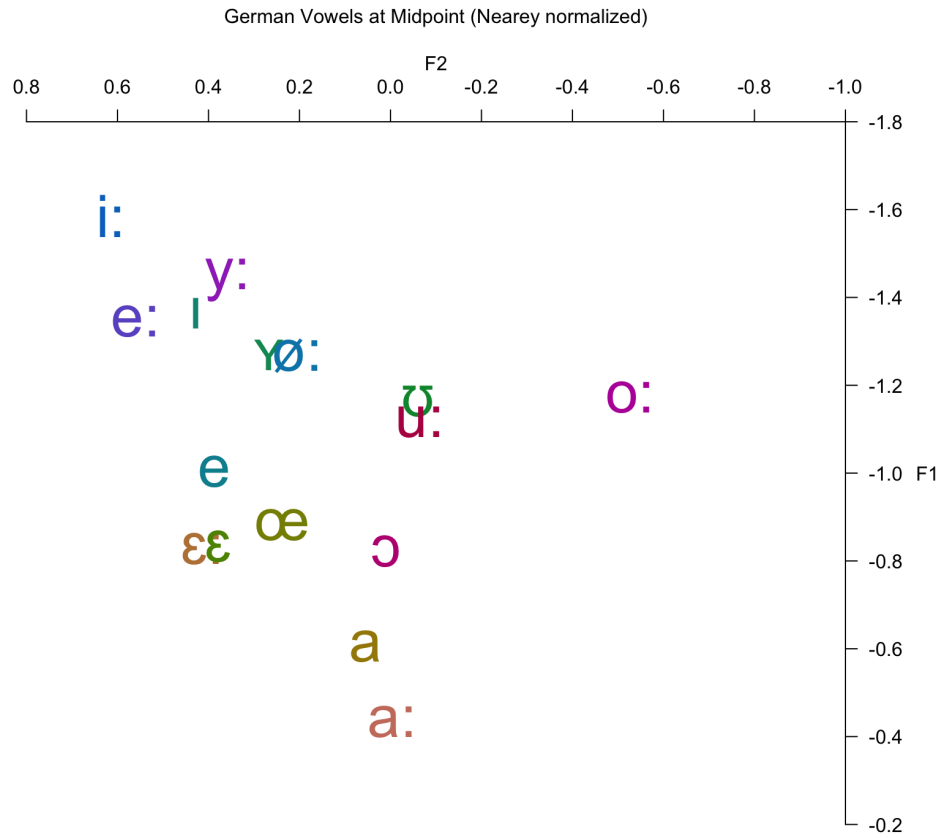


Fig. 13: German vowels (red) and American English vowels (blue)

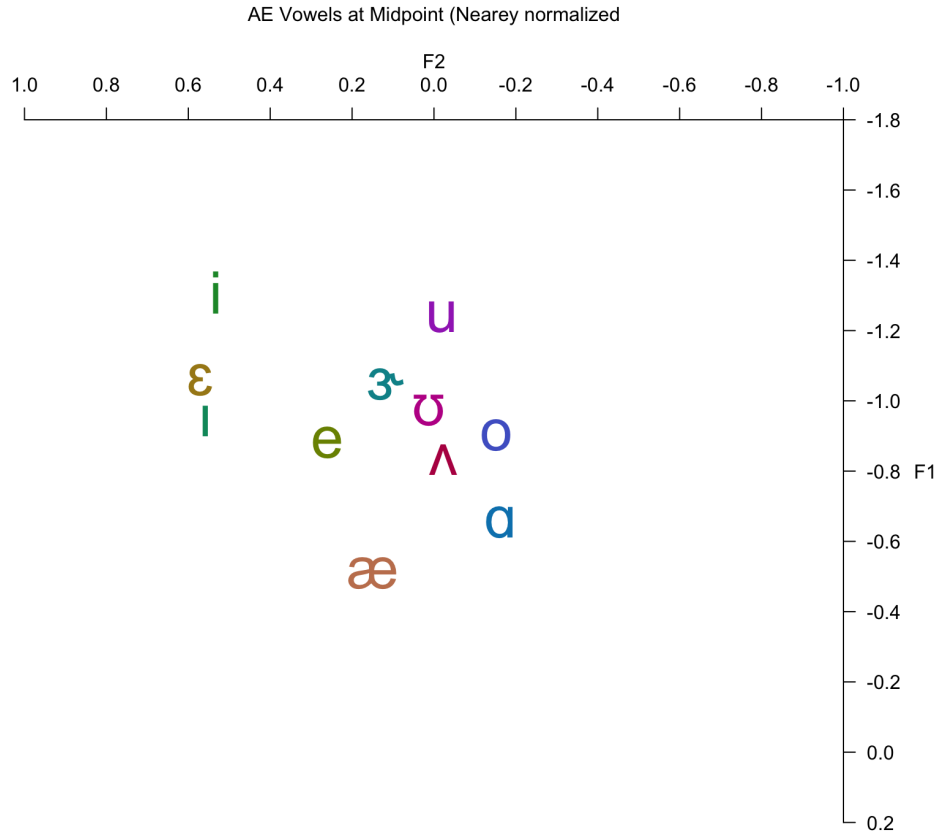
As mentioned previously, /u/, /ʊ/, and /o/ are fronted in California vowels, which causes them to move into the space of the German front rounded vowels (/y:/, /ʏ/, /ø:/, /œ/). This is reflected in the results, with German /y:/ and /ø:/ being perceived as American English /u/, and German /ʏ/ being perceived as American English /ʊ/.

2.2.7 Discussion

The results from experiment one show that while AE listeners relied mainly on spectral cues when confronted with novel sounds, they also utilize duration in some cases. Additionally, they also show flexibility in adapting their cue weighting to account for unknown sounds, such as the German short front rounded /ʏ/, /y:/, and /ø:/. One prediction for the English listeners was that duration as a cue would not be used and listeners would instead rely mainly on spectral features. This prediction was only true for some vowel pairs. However, duration was never used alone but rather in conjunction with spectral information as shown in the vowel-specific models. This indicates that AE listeners use duration only as a secondary cue. Figure 14 shows the mean values calculated from F1 and F2 midpoints of English and German vowels on the F1/F2 plane. German has more peripheral vowels than English, most likely due to the higher number of vowels in the vowel inventory.



(a)



(b)

Fig. 14: F1/F2 means at midpoints for (a) German and (b) American English vowels

When comparing vowel locations at the midpoint in the F1/F2 plane, we can see that listeners relied on F1 at midpoint heavily, for example, German /i:/ and /e:/ were both perceived as AE /i/, and German /ɛ:/, /ɛ/, and /e/ were perceived as AE /e/. Additionally, listeners seem to be relying on vowel height and frontness/backness of the vowels. German front high /e:/ and /i:/ for example were mapped to the AE front high /i/ with a goodness of fit rating of 7.

Another factor could be rounding, with the German high front rounded /y:/ and /ø:/ being mapped to English high back rounded /u/. The goodness of fit rating was 5 and 4 respectively, which is likely due to the fact that AE does not have high front

rounded vowels. These findings also support the predictions based on the PAM model. The /y:/-/ø:/ → /u/ case is an example of the single category case, where two L2 sounds assimilate into the same L1 category but neither are rated as a good fit. The SLM model assumes vowels deemed most dissimilar to native categories will eventually emerge as a new category instead of merging with an L1 category. The front rounded vowels also received lower goodness of fit ratings, which means listeners perceived them as less similar to their native vowel categories. Based on the goodness of fit ratings, listeners most likely recognized that these vowels were different and therefore used a different set of cues when trying to categorize the “worse” vowels.

Interestingly, almost all long German vowels’ goodness of fit ratings were higher than those of their short counterparts (with the exception of the long front rounded vowels /y:/, /ø:/, and /u:/). This could be triggered by AE vowels’ longer average duration (between 411ms and 570ms) in comparison with the short durations of German short vowels (between 263ms and 314ms), as shown in Figure 15. The long vowels were perceived as better fits and therefore might have been easier to perceive and categorize for AE listeners, since they fall into the duration range of AE vowels, while the short vowels differed on both spectral and durational dimensions. Taken together, these findings suggest that AE listeners do rely on spectral features as primary cues and duration only as a secondary cue, and not in all cases. Additionally, it seems that the goodness of fit judgments reflect knowledge about finely-grained phonetic details in the L1. Ratings of category goodness could also point to listeners’ awareness of variation in the acoustic target realization of vowels and what would be appropriate

allophonic variation in comparison to what examples of the non-native vowels would be inappropriate allophonic variations.

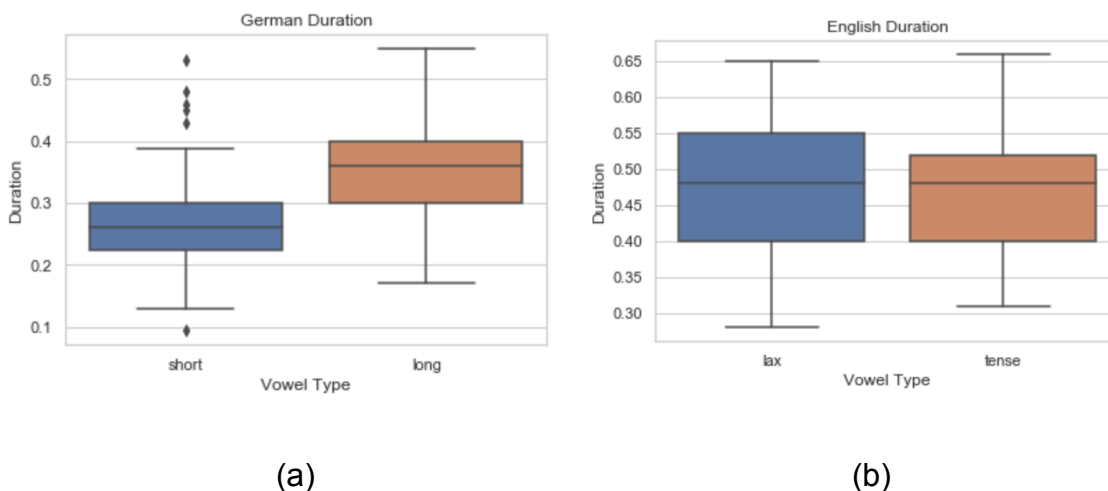


Fig. 15: Duration by vowel type. (a) Durations for German short and long vowels; (b) Durations for English lax and tense vowels

The findings from experiment one additionally provide evidence for language-specific speech perception processes that rely on a specific set of cues used to disambiguate vowels. Speech that deviates substantially from the patterns in the native language along multiple acoustic-phonetic dimensions seems to be recognized as “weird” and speakers adapt which cues they use to categorize these vowels in terms of their L1 vowel categories, even when the fit is rated not good. While AE listeners relied on spectral cues for identifying most categories, the “weird” vowels are of particular interest because they show that when confronted with novel sounds listeners adapt their cue weighting to identify these sounds.

Furthermore, the results from experiment one provide information about the acoustic realizations of the long-short distinction in German vowel pairs. While duration was significantly different between German long and short vowels, AE listeners were

able to pick up on spectral differences as well and exploit them in the identification and rating task. This means that while German vowels differ in quantity, they also differ in quality.

Whether native German listeners rely mainly on duration or on spectral information will be further investigated in experiments two and three. Additionally, the LDA run on corner vowels reached a higher accuracy when VISC information was excluded. Experiment four tests whether German listeners use VISC information in the perception of long/short vowels.

Chapter 3

3.1 Experiments two and three

The aim of experiments two and three is to examine the use of spectral and durational cues in the perception of German long/short vowels. Experiment two was designed to investigate the importance of duration in the discrimination of long and short vowel pairs using duration continua. Experiment three was designed to investigate the importance of spectral information in the discrimination of long and short vowel pairs using spectral continua.

3.2 Experiment two - Duration Continua

3.2.1 Introduction

Some of the literature on the perception of long versus short vowels in German claims that duration is the main cue for discriminating between the long and short pairs (c.f. von Essen 1979, Heike 1969, 1970, 1972, Lindner 1976, Weiss 1976, Sendlmeier 1981, Bennett 1968). To test this hypothesis, native German listeners were presented with vowel continua consisting of five different duration steps for every vowel. Both long and short vowels were manipulated in five steps using the original duration as measured in the data from experiment 1 as start (originally long duration) and end (originally short

duration) points. Formant frequencies were kept constant to either long formant frequencies (FFS) or short FFS in the five continua, resulting in 10 continua per long/short vowel pair. Listeners had to identify vowels as either long or short, using minimal pairs on the screen to decide whether a token was long or short.

If German listeners use duration as the primary cue in vowel identification, it should not matter whether a token contains originally long or short spectral information, and listeners will categorize shorter stimuli as short vowels, even if the token is spectrally based on a long vowel. Reversely, if spectral information is the primary cue, duration manipulations should not matter and vowels should still be identified as long or short based on the spectral pattern.

3.2.2 Methods and Materials

3.2.2.1 Listeners

65 native speakers of German were recruited via Linguist List for this study (m = 18, f = 44, non-binary = 3, mean age = 28.5, age σ = 10.7) and completed the experiment on Qualtrics. All participants reported German as their native language and all but ten reported being fluent in one or more other languages. None reported any problems with the experiment platform. A table with detailed demographic information can be found in the appendix.

3.2.2.2 Stimuli

The acoustic data from experiment one was subsetted to only include male measurements. For each vowel, the mean duration was calculated in Python. The means for each long and short vowel pair were used as start and end points for the duration continua. F1 and F2 were set to the average values calculated from the midpoints of the naturally produced speech in experiment one and kept constant for either long or short formant values within a duration continuum. The synthesized vowels were manipulated in five duration steps from start point 1 (long) to end point 5 (short) for each vowel. By synthesizing vowels and only manipulating the duration, the possibility of other acoustic cues interfering with identification can be excluded. The vowel pairs used were /i:/-/ɪ/, /y:/-/ʏ/, /u:/-/ʊ/, /ø:/-/œ/, /o:/-/ɔ/, /e:/-/ɛ/, /a:/-/a/, /ɛ:/-/ɛ/. This resulted in a total of 80 stimuli (8 vowels x 5 duration steps x two original FFS conditions).

3.2.2.3 Procedure

The experiment was conducted fully online, using the Qualtrics survey platform. Participants were instructed to sit in a quiet room and use a computer to complete the experiment. Before starting the trials, participants were able to play a test sentence and asked to set the volume to a comfortable level to ensure that their sound output worked and the sound was loud enough. Stimuli were presented in randomized order. In the experiment, subjects were presented with minimal pairs of real words on the screen and

had to choose between a long or short vowel response for each token. Listeners could not progress to the next trial if they had not chosen a response. Figure 16 shows an example screen.



Was haben Sie gehört?

Höhle

Hölle



Fig. 16: Trial screen for /ø:/-/œ/ with minimal pair Höhle - Hölle (cave - hell)

After completing the listening trials, participants were asked to fill out a demographic questionnaire. The experiment took an average of 10 minutes to complete.

3.2.3 Analysis

Responses were coded for whether the response was long (=1) or short (=0). The data were analyzed using a generalized mixed-effects logistic regression (*lme4* R package; Bates et al., 2015). Main effects included manipulation step (1, 2, 3 4, 5), formant frequency (manually coded to long = 1, short = 0), and their interaction. Random effects included by-Listener random intercepts and by-Vowel random intercepts (*glmer* syntax: `LongShortResponse ~ ManipulationStep * FFSLength + (1 | Vowel) + (1 | Listener_ID)`). To investigate whether these patterns hold true for all long/short vowel pairs, separate regression models were run for each pair using the same *glmer* syntax but only including by-Listener random intercepts.

3.2.4 Results

The output of the logistic regression model run on all vowels is provided in Table 14. Manipulation step and original FFS were significant main effects. The negative estimated coefficient for manipulation step indicates that listeners were less likely to select a token as long with rising manipulation step, Figure 17 shows this with listeners selecting tokens more frequently as short in the shorter continua steps.

In addition, there was an effect of FFS: while identifications for long tokens dropped overall as duration continua got shorter, tokens with originally long FFS were

still selected as long more often than those with originally short FFS. No other effects or interactions were observed.

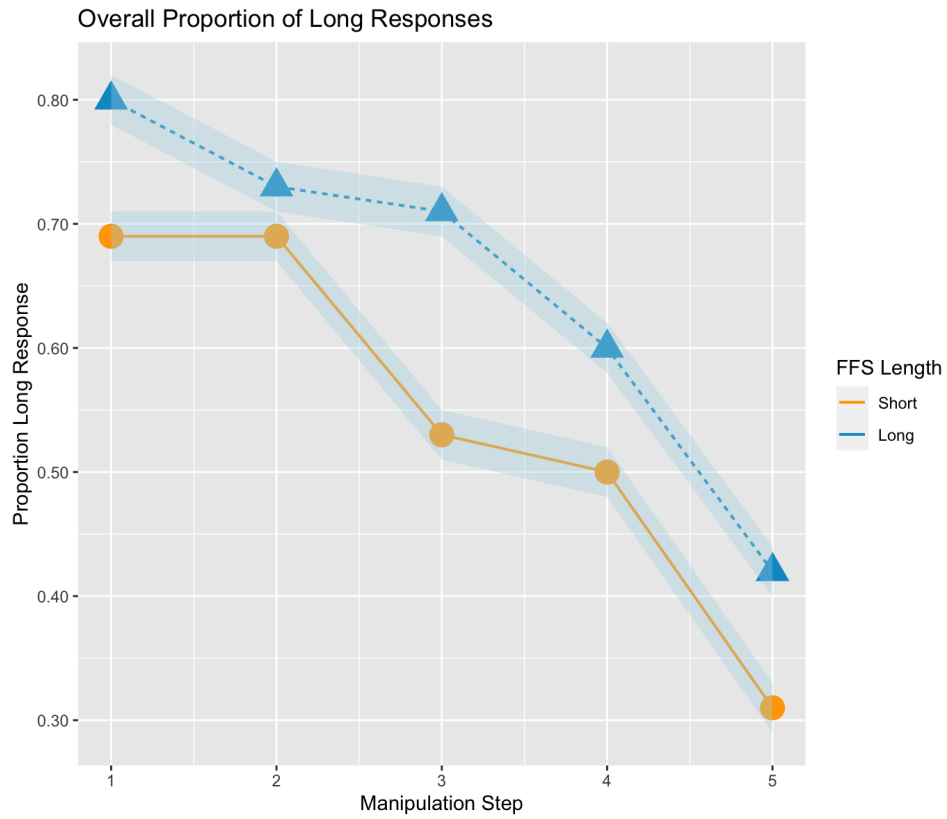


Fig. 17: Percentage of Responses selected as long for each manipulation step from 1 (originally long duration) to 5 (originally short duration) by formant frequencies with 95% confidence intervals

Table 14: Regression output for general duration manipulation model

	<i>Est</i>	<i>Std. Err.</i>	<i>z</i>	<i>p</i>
(Intercept)	1.74	0.35	4.94	<0.001**
ManipulationStep	-0.50	0.03	-14.96	<0.001**
FFSLength	0.59	0.16	3.63	0.001 ***
ManipulationStep:FFSLength	0.001	0.04	0.033	0.97

To investigate whether this pattern held true for all vowels, vowel specific analyses were run.

For /a:/-/a/ the same main effects were significant, but additionally, there was a significant interaction between manipulation step and original FFS length. Table 15 shows the regression output. The interaction was more closely examined with Tukey’s HSD pairwise comparisons within the model using the *emmeans()* function in the *emmeans* R package (Lenth et al., 2021). This revealed that listeners selected tokens with originally short FFS significantly less as long for the third manipulation step than tokens that contained originally long FFS ($p < .0001$). Figure 18 shows the percentage of long responses selected for /a:/-/a/.

Table 15: Regression output for /a:/-/a/

	<i>Est</i>	<i>Std. Err.</i>	<i>z</i>	<i>p</i>
(Intercept)	2.31	0.34	6.73	<0.001**
ManipulationStep	-0.75	0.10	-7.38	<0.001**
FFSLength	3.67	0.77	4.72	<0.001*
ManipulationStep:FFSLength	-0.48	0.19	-2.51	0.012*

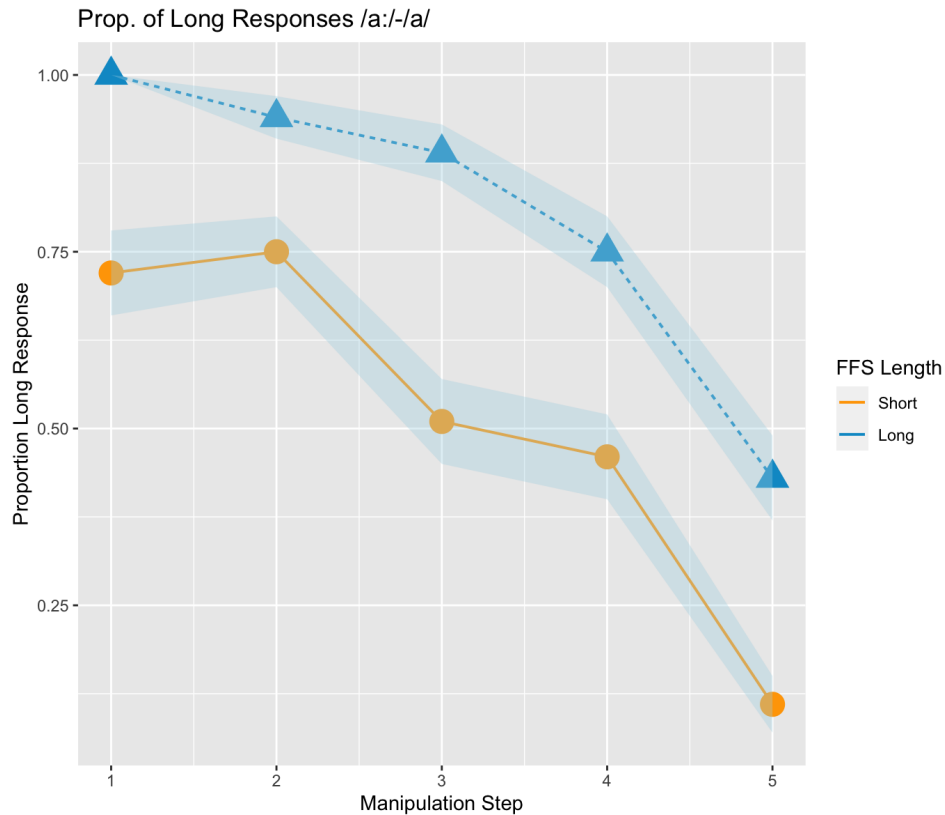


Fig. 18: Percentage of Responses selected as long for each manipulation step from 1 (originally long duration) to 5 (originally short duration) by formant frequencies for /a:/-/a/ with 95% confidence intervals

For /ε:/-/ε/ only manipulation step was a significant main effects. No other significant effects were observed. Table 16 shows the regression output. Figure 19 shows the percentage of long responses selected for /ε:/-/ε:/. While manipulation step was a significant predictors of whether the vowels were perceived as long or short, /ε:/-/ε/ was perceived as long the majority of the time with over 70% long responses even as continua approached the short durations. This shows a listeners bias towards /ε:/.

Table 16: Regression output for /ɛ:/-/ɛ/

	<i>Est</i>	<i>Std. Err.</i>	<i>z</i>	<i>p</i>
(Intercept)	3.46	0.51	6.81	<0.001**
ManipulationStep	-0.44	0.12	-3.48	<0.001**
FFSLength	0.007	0.68	0.01	0.99
ManipulationStep:FFSLength	0.03	0.18	0.17	0.86

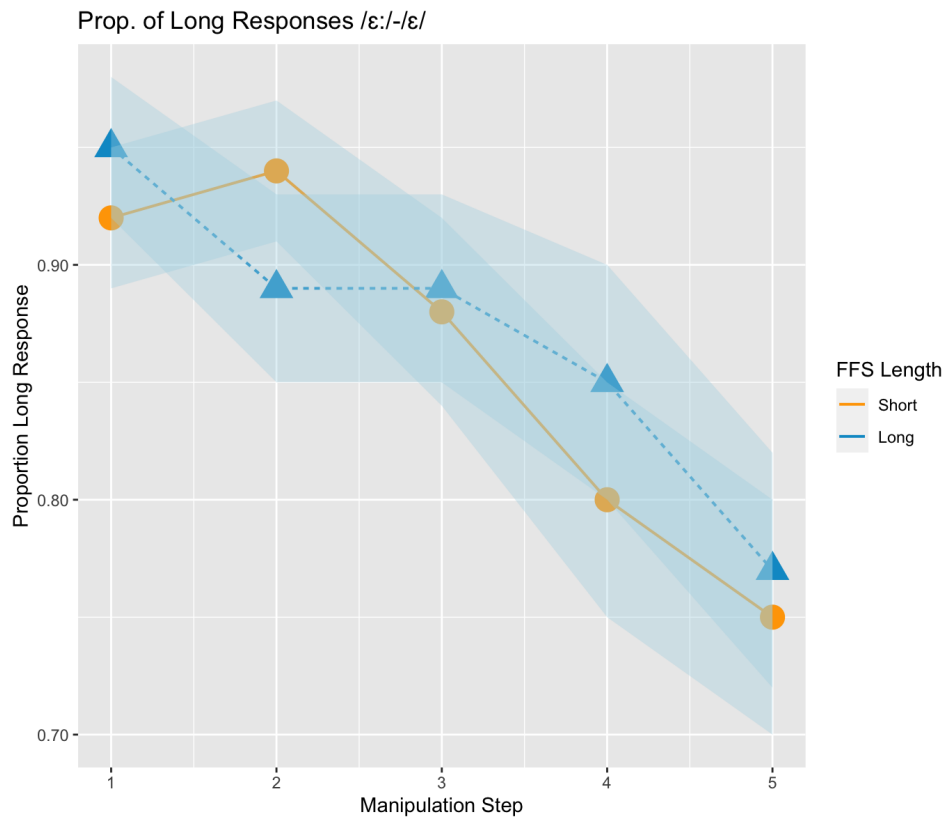


Fig. 19: Percentage of Responses selected as long for each manipulation step from 1 (originally long duration) to 5 (originally short duration) by formant frequencies for /ɛ:/-/ɛ:/ with 95% confidence intervals

For /e:/-/e/ manipulation step and original FFS were significant main effects, additionally there was a significant interaction between manipulation step and original FFS length. Table 17 shows the regression output. The interaction was more closely

examined with Tukey's HSD pairwise comparisons within the model using the *emmeans()* function in the *emmeans* R package (Lenth et al., 2021). This revealed that listeners selected tokens with originally short FFS significantly less as long for the third manipulation step than tokens that contained originally long FFS ($p < .0001$). Figure 20 shows the percentage of long responses selected for /e:/-/e/.

Table 17: Regression output for /e:/-/e/

	<i>Est</i>	<i>Std. Err.</i>	<i>z</i>	<i>p</i>
(Intercept)	2.72	0.39	6.83	<0.001**
ManipulationStep	-0.57	0.10	-5.53	<0.001**
FFSLength	3.87	0.95	4.05	<0.001*
ManipulationStep:FFSLength	-0.59	0.22	-2.62	0.008*

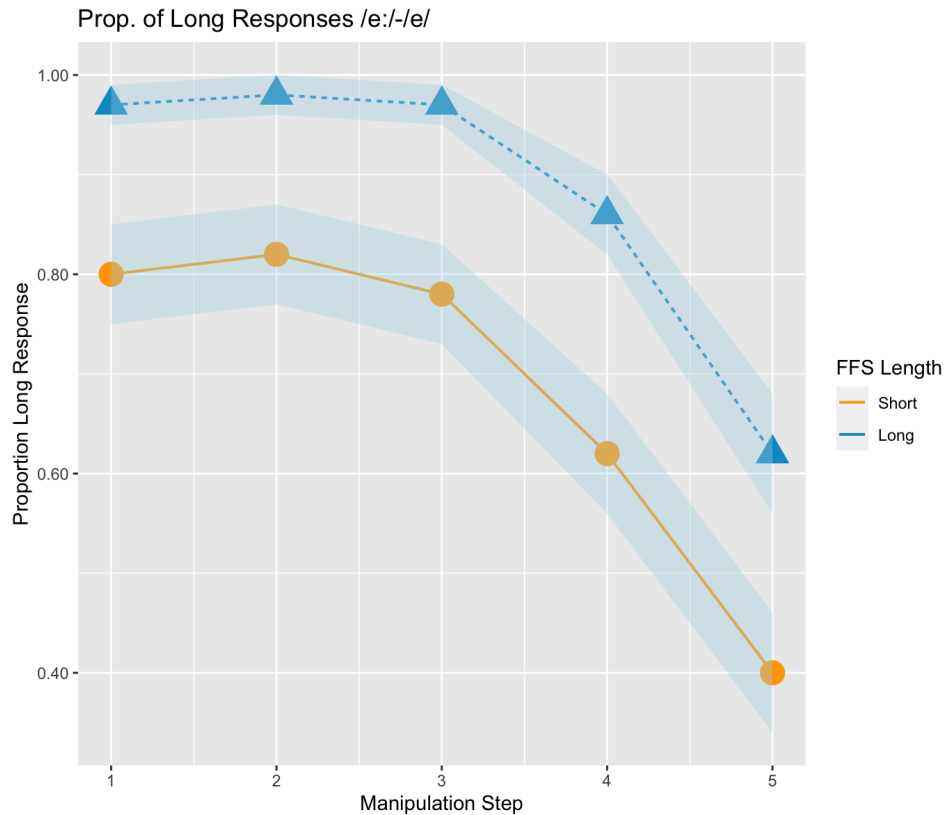


Fig. 20: Percentage of Responses selected as long for each manipulation step from 1 (originally long duration) to 5 (originally short duration) by formant frequencies for /e:/-/e/ with 95% confidence intervals

For /i:/-/ɪ/ manipulation step and original FFS were significant main effects. No other significant effects were observed. Table 18 shows the regression output. Figure 21 shows the percentage of long responses selected for /i:/-/ɪ/, in which tokens containing originally long FFS were selected as long *less* frequently overall than those containing originally short FFS. This is also reflected in the negative estimated coefficient for FFS length. This could be due to the synthesized vowels not sounding natural for this vowel pair, as /i:/ has a very high F2 and is the most peripheral vowel in the German vowel space on the F2 scale.

Table 18: Regression output for /i:/-/l/

	<i>Est</i>	<i>Std. Err.</i>	<i>z</i>	<i>p</i>
(Intercept)	1.67	0.31	5.32	<0.001**
ManipulationStep	-0.51	0.09	-5.66	<0.001**
FFSLength	-2.12	0.43	-4.91	<0.001*
ManipulationStep:FFSLength	0.22	0.13	1.70	0.08

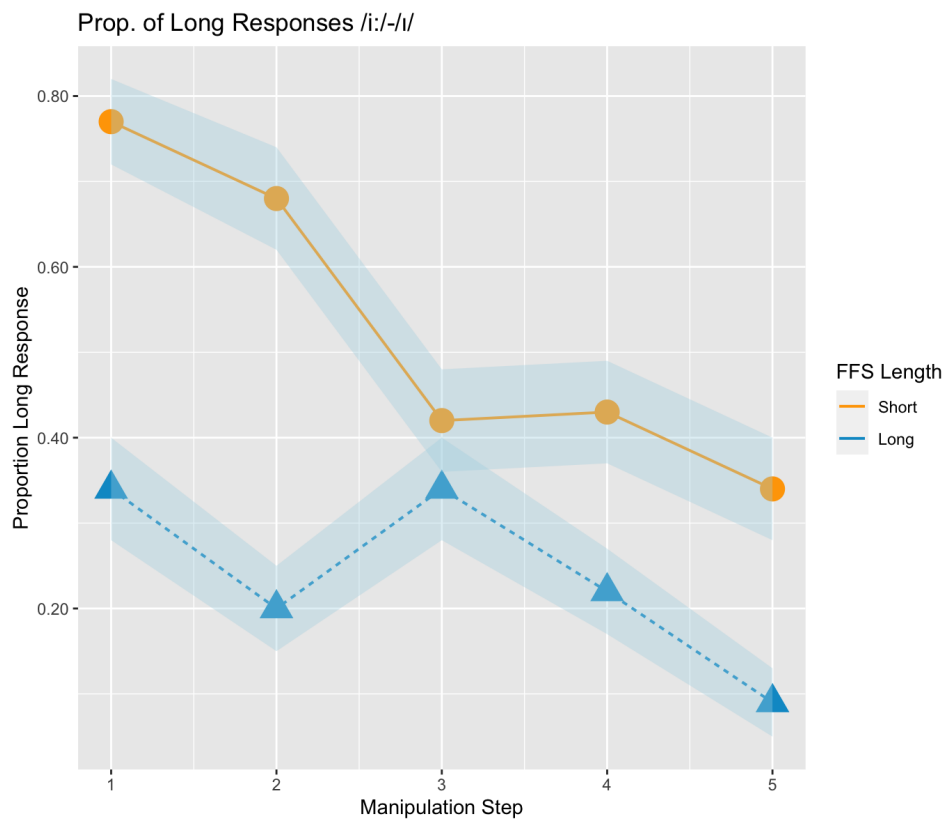


Fig. 21: Percentage of Responses selected as long for each manipulation step from 1 (originally long duration) to 5 (originally short duration) by formant frequencies for /i:/-/l/ with 95% confidence intervals

For /o:/-/ɔ/ manipulation step and original FFS were significant main effects. No other significant effects were observed. Table 19 shows the regression output. Figure 22 shows the percentage of long responses selected for /o:/-/ɔ/.

Table 19: Regression output for /o:/-/ɔ/

	<i>Est</i>	<i>Std. Err.</i>	<i>z</i>	<i>p</i>
(Intercept)	3.32	0.43	7.68	<0.001**
ManipulationStep	-0.74	0.11	-6.67	<0.001**
FFSLength	0.59	0.57	1.02	0.30
ManipulationStep:FFSLength	-0.17	0.15	-1.09	0.27

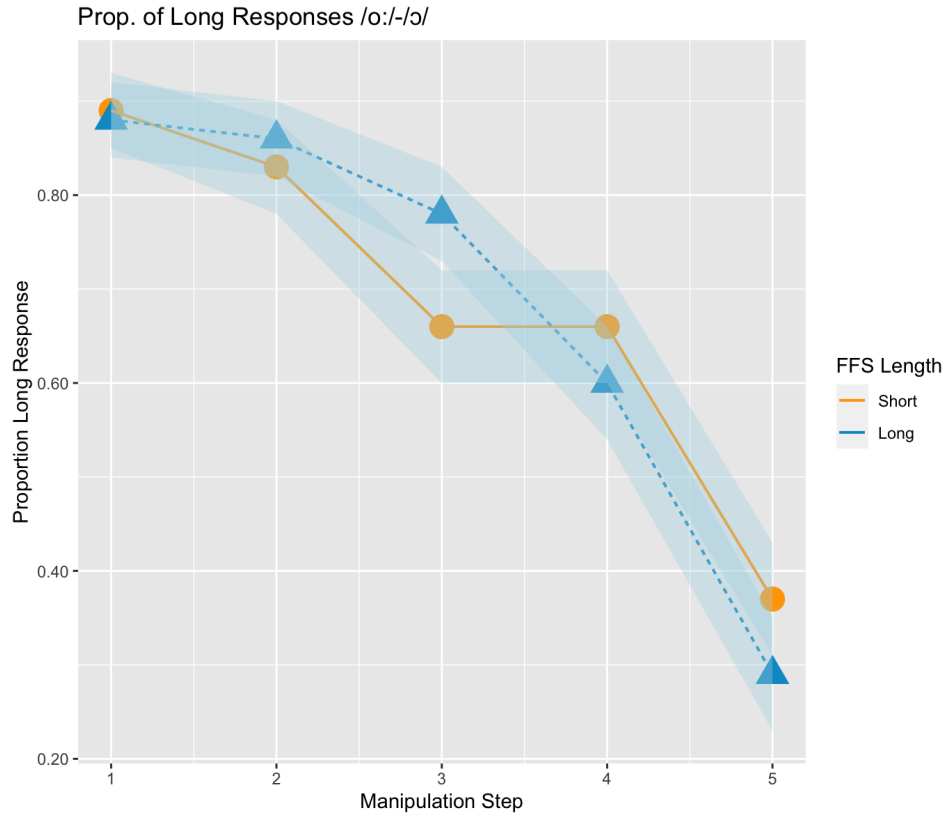


Fig. 22: Percentage of Responses selected as long (blue) or short (orange) for each manipulation step from 1 (originally long duration) to 5 (originally short duration) by formant frequencies for /o:/-/ɔ/ with 95% confidence intervals

For /ø:/-/œ/ manipulation step and original FFS were significant main effects. No other significant effects were observed. Table 20 shows the regression output. Figure 23 shows the percentage of long responses selected for /ø:/-/œ/.

Table 20: Regression output for /ø:/-/œ/

	<i>Est</i>	<i>Std. Err.</i>	<i>z</i>	<i>p</i>
(Intercept)	1.03	0.29	3.44	<0.001**
ManipulationStep	-0.71	0.11	-6.58	<0.001**

FFSLength	2.99	0.56	5.34	<0.001*
ManipulationStep:FFSLength	-0.16	0.16	-1.04	0.29

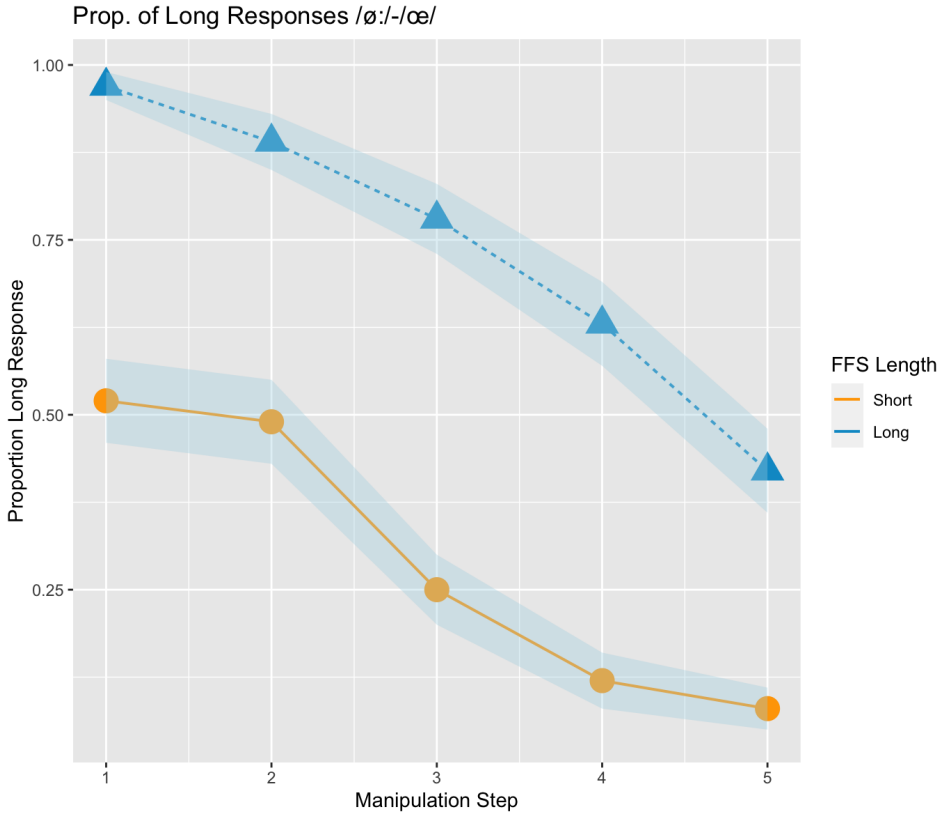


Fig. 23: Percentage of Responses selected as long (blue) or short (orange) for each manipulation step from 1 (originally long duration) to 5 (originally short duration) by formant frequencies for /ø:/-/œ/ with 95% confidence intervals

For /y:/-/ʏ/ manipulation step and original FFS were significant main effects. No other significant effects were observed. Table 21 shows the regression output. Figure 24 shows the percentage of long responses selected for /y:/-/ʏ/.

Table 21: Regression output for /y:/-/ʏ/

	<i>Est</i>	<i>Std. Err.</i>	<i>z</i>	<i>p</i>
(Intercept)	0.86	0.31	2.81	0.004*
ManipulationStep	-0.44	0.09	-4.80	<0.001**
FFSLength	1.82	0.46	3.88	<0.001*
ManipulationStep:FFSLength	-0.07	0.13	-0.56	0.57

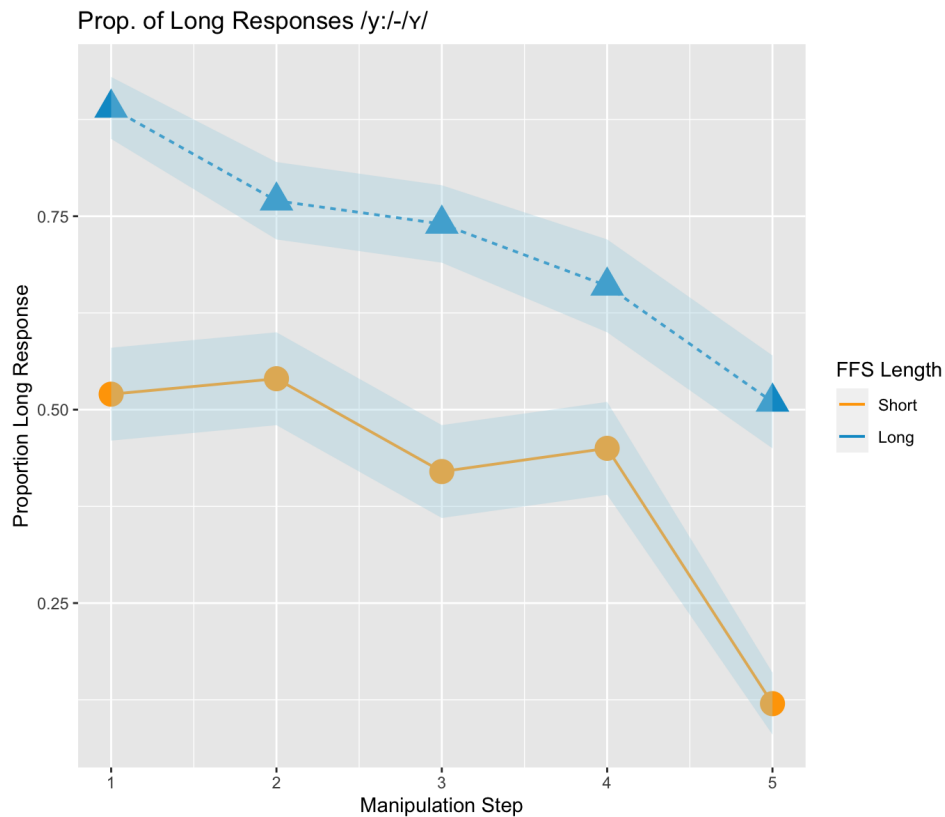


Fig. 24: Percentage of Responses selected as long (blue) or short (orange) for each manipulation step from 1 (originally long duration) to 5 (originally short duration) by formant frequencies for /y:/-/ʏ/ with 95% confidence intervals

In the case of /u:/-/ʊ/ no significant main effects were found. As mentioned in section 2.2.6.1.1 in chapter two, this is likely due to a flaw in the experimental design where the spellings to record these two vowels were not clearly indicating a short vowel

(*But* instead of *Butt*). The synthesized vowels were based on the natural recordings from experiment one, which caused this flaw to translate into the second and third experiments.

3.2.5 Interim Discussion

The results from experiment two provide evidence for both quantity and quality being important cues in distinguishing between long and short vowel pairs in native German speakers. With continua approaching the short durations, listeners were less likely to select them as long. This pattern is repeated in the vowel-specific analyses, which show a negative estimated coefficient for manipulation step for all vowels. However, instead of observing a steep categorical boundary, the overall pattern shown in figure MAINEFFECSDUR is gradual. This is no surprise because native listeners also utilized FFS as a cue, in all cases but /o:/-/ɔ/ and /ɛ:/-/ɛ/, which was kept steady within the continua to be either originally long or originally short. Native listeners seem to exploit these differences when quantity alone is not sufficient to make the distinction between a long or a short vowel. This listener behavior points to FFS being used as a secondary cue in identifying whether a vowel is long or short in German. Experiment three will investigate the role of spectral information more closely.

3.3 Experiment three - Spectral Continua

Some of the literature on the perception of long versus short vowels in German claims that quality is the main cue for discriminating between the long and short pairs (c.f. Bennett 1968, Ungeheuer 1969, Strange and Bohn 1998). Experiment three was designed to test this hypothesis. Native German listeners were presented with vowel continua consisting of five different quality steps for every vowel pair. Both long and short vowels were manipulated in five steps using the original FFS as measured in the data from experiment 1 as start (originally short FFS) and end (originally long FFS) points. Duration was kept constant to either long duration or short duration in the five continua, resulting in 10 continua per long/short vowel pair. Listeners had to identify vowels as either long or short, using minimal pairs on the screen to decide whether a token was long or short.

If German listeners use quality as the primary cue in vowel identification, it should not matter whether a token contains an originally long or short duration and listeners will categorize stimuli containing short FFS as short vowels, even if the token has an originally long duration. Reversely, if duration is the primary cue, FFS manipulations should not matter and vowels should still be identified as long or short based on the duration pattern.

3.3.1 Methods and Materials

3.3.1.1 Listeners

57 native speakers of German were recruited via Linguist List for this study (m = 19, f = 36, non-binary = 2, mean age = 36.3, age σ = 14.6) and completed the experiment on Qualtrics. All participants reported German as their native language and all but sixteen reported being fluent in one or more other languages. None reported any problems with the experiment platform. A table with detailed demographic information can be found in the appendix.

3.3.1.2 Stimuli

The acoustic data from experiment one was subsetted to only include male measurements. The reason for this is that a male voice sounds more natural when synthesizing individual vowels with phonTools (Barreda 2015) in R. For each vowel, the means for F1, F2, and F3 were calculated in Python. The means for each long and short vowel pair were used as start and end points for the spectral continua. Duration was set to the average values calculated from the midpoints of the naturally produced speech in experiment one and kept constant for either long or short duration values within a spectral continuum. The synthesized vowels were manipulated in five spectral steps from start point 1 (short) to end point 5 (long) for each vowel. By synthesizing vowels and only manipulating the spectral information, the possibility of other acoustic

cues interfering with identification can be excluded. The vowel pairs used were /i:/-/ɪ/, /y:/-/ʏ/, /u:/-/ʊ/, /ø:/-/œ/, /o:/-/ɔ/, /e:/-/ɛ/, /a:/-/a:/. This resulted in a total of 80 stimuli.

3.3.1.3 Procedure

The experiment was conducted fully online, using the Qualtrics survey platform. Participants were instructed to sit in a quiet room and use a computer to complete the experiment. Before starting the trials, participants were able to play a test sentence and asked to set the volume to a comfortable level to ensure that their sound output worked and the sound was loud enough. Stimuli were presented in randomized order. In the experiment, subjects were presented with minimal pairs of real words on the screen and had to choose between a long or short vowel response for each token. Listeners could not progress to the next trial if they had not chosen a response. Figure 25 shows an example screen.

Was haben Sie gehört?

Höhle

Hölle



Fig. 25: Trial screen for /ø:/-/œ/ with minimal pair Höhle - Hölle (cave - hell)

After completing the listening trials, participants were asked to fill out a demographic questionnaire. The experiment took an average of 10 minutes to complete.

3.3.2 Analysis

Responses were coded for whether the response was long (=1) or short (=0). The data were analyzed using a generalized mixed-effects logistic regression (*lme4* R package; Bates et al., 2015). Main effects included manipulation step (1, 2, 3, 4, 5), duration (manually coded as long = 1, short = 0), and their interaction. Random effects included by-Listener random intercepts and by-Vowel random intercepts (*lmer* syntax: `LongShortResponse ~ ManipulationStep * DurLength + (1 | Vowel) + (1 | Listener_ID)`).

To investigate whether these patterns hold true for all long/short vowel pairs, separate regression models were run for each pair using the same *glmer* syntax but only including by-Listener random intercepts.

3.3.3 Results

The output of the logistic regression model is provided in Table REGSPEC. As expected, original duration length was a significant main effect: as seen in Figure 26, listeners selected the originally durationally short tokens as short and the originally durationally long tokens as long regardless of the spectral manipulation step. Additionally, there was a significant interaction between manipulation step and original duration length. Table 22 shows the regression output.

The interaction was more closely examined with Tukey's HSD pairwise comparisons within the model using the *emmeans()* function in the *emmeans* R package (Lenth et al., 2021). This revealed that listeners selected tokens with an originally short duration significantly less as long for the third manipulation step than tokens that contained an originally long duration ($p < .0001$). Figure 26 shows the percentage of long responses selected for each manipulation step. No other effects were observed.

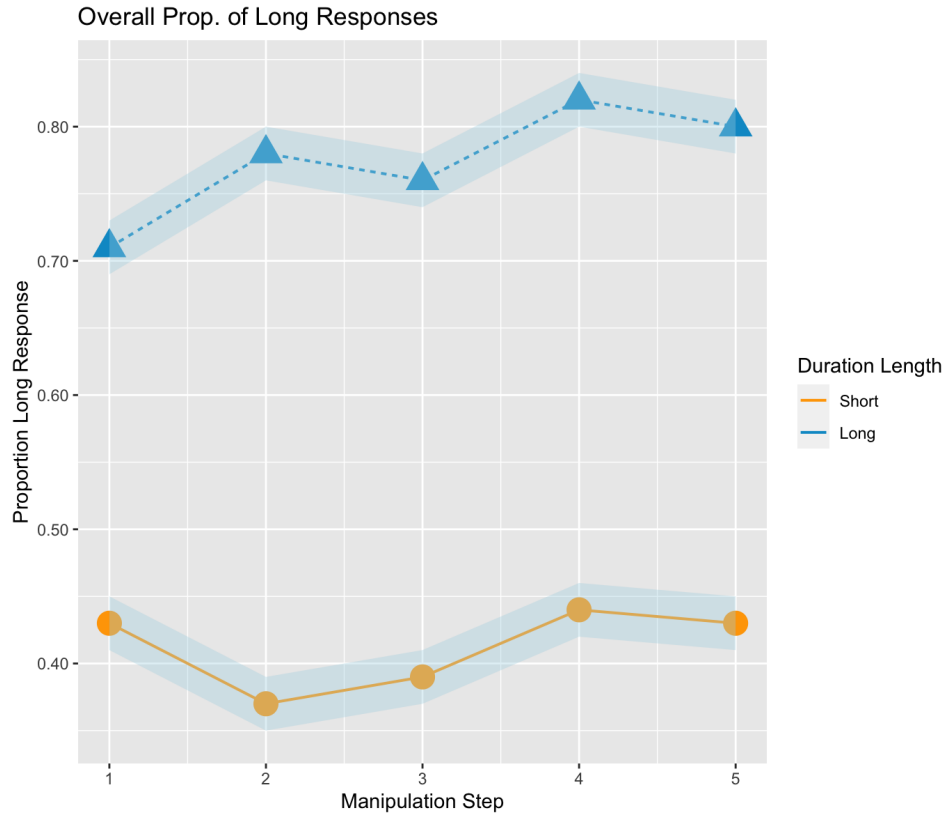


Fig. 26: Percentage of Responses selected as long each manipulation step from 1 (originally short FFS) to 5 (originally long FFS) by duration with 95% confidence interval

Table 22: Regression output for general spectral manipulation model

	<i>Est</i>	<i>Std. Err.</i>	<i>z</i>	<i>p</i>
(Intercept)	-0.51	0.31	-1.61	0.10
ManipulationStep	0.03	0.03	0.90	0.36
DurLength	1.58	0.16	9.48	<0.001 ***
ManipulationStep:DurLength	0.11	0.05	2.28	0.02*

To investigate whether this pattern held true for all vowels, vowel specific analyses were run. With the exception of /œ/-/ø:/, /ʊ/-/u:/, and /y/-/y:/ the same pattern of main effects were observed in all vowel specific analyses.

For /œ/-/ø:/, manipulation step was a significant main effect. Additionally, a significant interaction between manipulation step and original duration was observed, as shown in Table 23 and figure 27. As formant manipulations approached the originally long FFS listeners were more likely to select the token as long in both original duration conditions. The interaction was more closely examined with Tukey's HSD pairwise comparisons within the model using the *emmeans()* function in the *emmeans* R package (Lenth et al., 2021). This revealed that listeners selected tokens with an originally short duration significantly less as long for the third manipulation step than tokens that contained originally long duration ($p < .0001$). Figure 27 shows the percentage of long responses selected for each manipulation step. No other effects were observed.

Table 23: Regression output for /œ/-/ø:/

	<i>Est</i>	<i>Std. Err.</i>	<i>z</i>	<i>p</i>
(Intercept)	-0.007	0.31	-0.02	0.97
ManipulationStep	-0.29	0.09	-3.04	0.002**
DurLength	-0.53	0.45	-1.17	0.23
ManipulationStep:DurLength	1.00	0.16	6.21	<0.001**

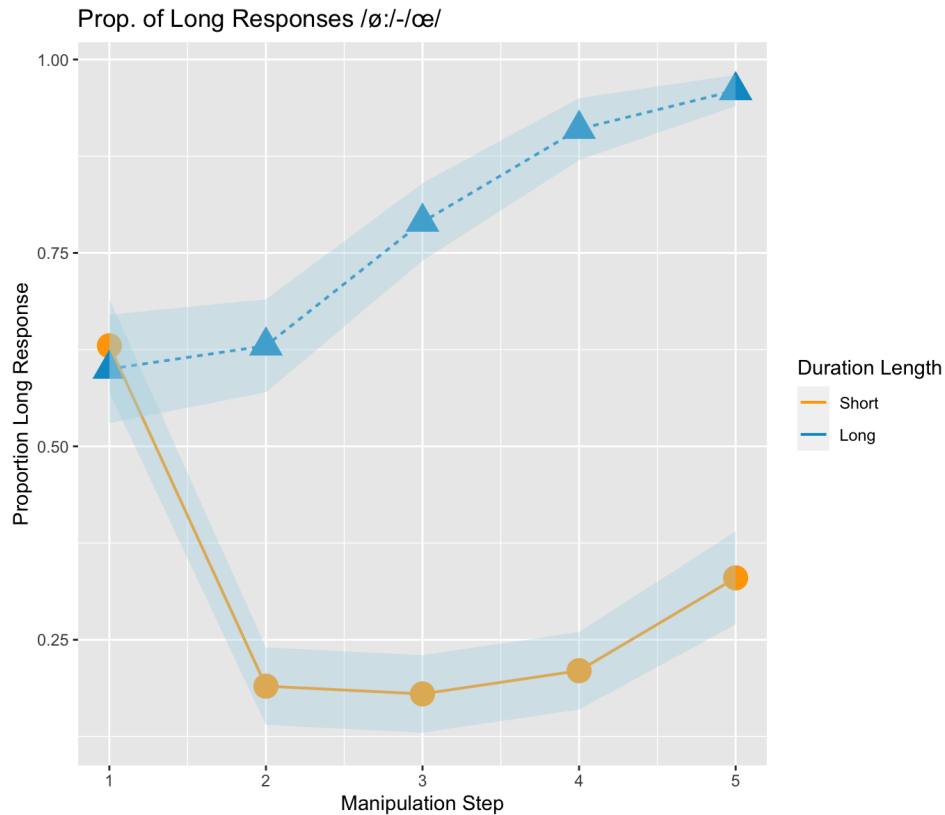


Fig. 27: Percentage of Responses selected as long for each manipulation step from 1 (originally short FFS) to 5 (originally long FFS) by duration for /œ/-/ø:/

For /u:/-/ʊ/ only manipulation step was a significant main effect, as shown in Table 24 and Figure 28. No other effects or interactions were observed. With manipulation steps approaching the originally long FFS listeners were *less* likely to select the token as long, regardless of original duration. It should be noted that overall, listeners were below chance for all conditions, except the second manipulation step for originally long /u:/. As mentioned previously, this is likely due to a design flaw in the recordings of /u:/-/ʊ/, with short /ʊ/ being produced between /u:/ and a true production of /ʊ/. For more information on this see section 2.2.6.1.1 in chapter two.

Table 24: Regression output for /u:/-/ʊ/

	<i>Est</i>	<i>Std. Err.</i>	<i>z</i>	<i>p</i>
(Intercept)	-0.31	0.40	-0.78	0.43
ManipulationStep	-0.22	0.10	-2.11	0.03*
DurLength	-0.17	0.47	-0.36	0.71
ManipulationStep:DurLength	0.18	0.14	1.3	0.19

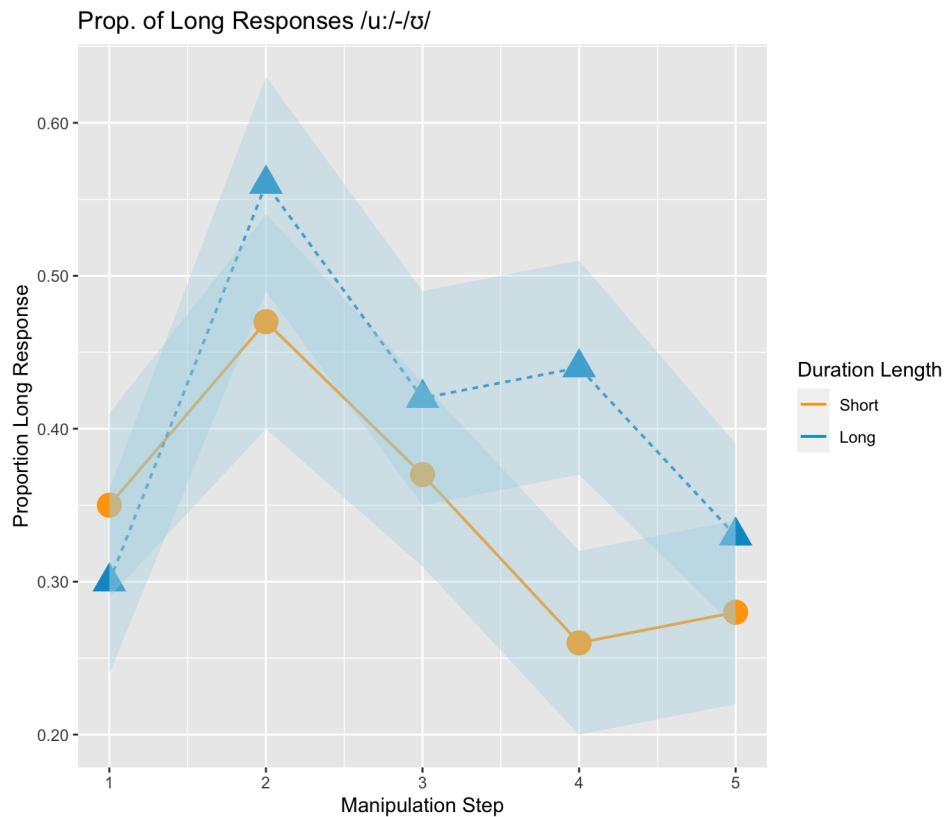


Fig. 28: Proportion of Responses selected as long for each manipulation step from 1 (originally short FFS) to 5 (originally long FFS) by duration for /u:/-/ʊ/

For /y:/-/ɣ/ original duration and manipulation step were significant, as shown in Table 25 and Figure 29. No other significant interactions were observed. With manipulation steps approaching originally long FFS listeners were more likely to select them as long even when the token contained an originally short duration.

Table 25: Regression output for /y:/-/Y/

	<i>Est</i>	<i>Std. Err.</i>	<i>z</i>	<i>p</i>
(Intercept)	-1.009	0.33	-2.99	0.002**
ManipulationStep	0.24	0.09	2.63	0.008**
DurLength	1.18	0.46	2.54	0.01*
ManipulationStep:DurLength	0.22	0.14	1.5	0.13

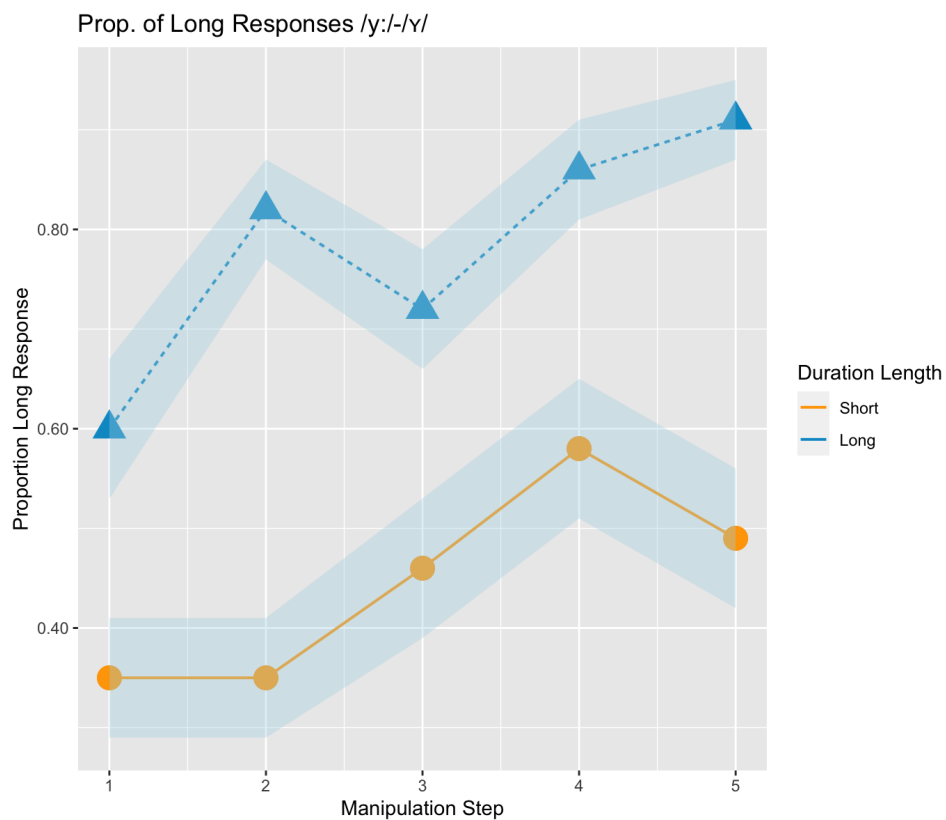


Fig. 29: Proportion of Responses selected as long for each manipulation step from 1 (originally short FFS) to 5 (originally long FFS) by duration for /y:/-/Y/

3.3.4 Interim Discussion

The results from experiment three confirm the results from experiment two in that quantity is used as the primary cue in distinguishing between long and short vowel pairs in German overall. However, we also see more vowel-specific patterns in experiment three. Recall that Weiss (1976) suggested that the importance of spectral information versus duration information was a function of vowel height with listeners relying more on duration to identify low vowels, but using spectral information to identify high vowels. In the third experiment, this pattern explains the results seen for /y:/-/ʏ/ and /ø:/-/œ/, for both pairs formant manipulation step was a significant predictor of whether a vowel was perceived as long or short. Therefore, experiment three provides more evidence for vowel-specific perception patterns. The vowel specific patterns also provide more information about how sensitive listeners are to changes in FFS. The FFS manipulations show some confusion for the front rounded vowels, instead of steadily selecting vowels as long more often as continua approached the long FFS, listeners selected manipulation step 2 as long more often than manipulation step 3 in the /y:/-/ʏ/ continua when the original duration was long, and step 4 as long more often than step 5, when the original duration was short. For /ø:/-/œ/, continua made with long *and* short durations were perceived as long approximately 60% of the time in the first manipulation step and only started to diverge after the first step. Conclusively, German listeners might rely on spectral targets being articulated with a high degree of precision to differentiate them from back rounded vowels. Harrington et al. (2011) found that languages that contrast high front rounded and high back rounded vowels, such as German, show a

greater magnitude and velocity of tongue dorsum retraction in order to ensure that the high back rounded vowel it is produced acoustically distinct from the high front rounded vowel.

Overall, the perception data point to FFS being used as a secondary cue in identifying whether a vowel is long or short in German when quantity alone is not sufficient. Additionally, while there was an effect for /u:/-/ʊ/, it was not in the expected direction with listeners selecting token as long less frequently as continua approached originally long FFS regardless of original duration. Recall that in experiment two there were no observed main effects for /u:/-/ʊ/. The observed pattern could be due to a flaw in the experimental design from experiment one, where the orthographic representation of the target word did not clearly indicate a short vowel (*But* instead of *Butt*), as mentioned above. This could have resulted in the short vowel /ʊ/ being acoustically realized somewhere between the true short vowel and the long vowel. Additionally, in the production data, both /u:/ and /ʊ/ showed more variation in F2 ($\sigma = 0.58$ and $\sigma = 0.27$ respectively) than the other vowels. This could lead to listeners deeming F2 to be an unreliable cue and less salient than duration and FFS continua being interpreted in an unreliable way. However, since there was also no observed effect for duration in either experiment, the stimuli might not have been informative because of the synthesis method. Additionally, the pattern seen for the /u:/-/ʊ/ pair could be due to high back rounded vowels being unstable in the production of FFS and often moving into the space of the high front rounded vowels (Hoole and Kühnert 1995, Harrington et al. 2011), therefore rendering spectral cues unreliable as well. Similarly, Tomaschek et al.

(2015) have shown that listeners were insensitive to spectral changes in the /u:/-/ʊ/ space.

3.4 General Discussion

The aim of experiments two and three was to investigate the relative perceptual importance of quantity and quality cues in the distinction of long and short vowels in German. Cue weighting was explored by using duration continua where FFS were kept constant, and spectral continua where duration was kept constant.

Results show that German listeners rely mainly on duration in the perception of long versus short vowels. While listeners did use FFS as a secondary cue in experiment two, in that tokens containing originally long FFS were more often selected as long than those containing short FFS, the proportion of tokens selected as long dropped overall as continua approached the short duration. Experiment three confirmed this result with tokens containing originally short durations being selected as long significantly less often over all FFS manipulation steps. While German long-short vowel pairs differ both spectrally and durationally in production, German listeners seem to rely mainly on duration differences to make the distinction between whether a vowel was long or short.

However, there was a clear effect of FFS being used as a secondary cue. If listeners used *only* quantity and ignored quality, the expected behavior for the continua would show no difference based on whether the FFS were long or short. Instead, the data show a clear difference between the originally long FFS tokens and the originally

short FFS tokens, in that listeners selected tokens containing originally long FFS more frequently as long even with decreasing durations.

This held true for all vowel pairs except /o:/-/ɔ/ and /ɛ:/-/ɛ/, in which listeners did not use original FFS at all. For /ɛ:/-/ɛ/, this could likely be due to the two vowels not contrasting as much in F1 as the other vowels pairs. Vowel height, acoustically represented by F1, has been shown to have a larger impact on vowel identification than F2 (Di Benedetto 1989). Similarly, the lack of effect for FFS in the perception of /o:/-/ɔ/ could be explained by the backness of the vowel pair. The German vowel space is a lot more crowded in the front than in the back as shown in figure VOWCHART. The only other back vowel pair is /u:/-/ʊ/. Additionally, /o:/ and /ɔ/ contrast in height, with /o:/ being a high vowel and /ɔ/ being a mid vowel. This could mean that listeners do not have to use secondary cues because the spectral separation of the back vowels is sufficient enough that listeners can rely on quantity to identify whether the vowel they heard was /o:/ or /ɔ/.

Overall, the results from experiment three suggest a more complex pattern of perception of length in German, where listeners are integrating both quantity and quality. Recall that earlier research stated that either only quality or only quantity is distinctive while the other one is redundant (c.f. Riad 1995, Wiese 1996, Lahiri and Dresher 1999, Vennemann 2000, Mangold 1990, Delattre 1969, Vernon 1976, Maack 1951, 1954, Jessen 1993, Weiss 1977). The data from experiments two and three show instead that while German listeners use duration as the primary cue, spectral information is not ignored and instead used as a secondary cue. Experiment four will

investigate the role of spectral information more closely by looking at the role of VISC in identifying whether a vowel is long or short in German.

Chapter 4

4.1 Introduction

Vowel inherent spectral change (VISC) refers to the change of formants over time or formant movement from the start point of a vowel to the end point. The term was coined by Nearey and Assmann (1986) and refers to the spectral changes associated with the vowel and not with the consonantal context in which a vowel occurs. Following Nearey's Compound Target Theory (1989), vowels can be differentiated using two points: the target and the offglide. This type of information is also referred to as dynamic in comparison to static, starting with Strange et al. (1976). In two studies, they found that listeners identified vowel targets more accurately in context (CVC syllables) than in isolation (V), and even when the middle portion of the vowel was spliced out of the CVC syllables and only the transitions were left, listeners were still able to identify the vowels. Later studies have also shown that VISC information leads to greater identification accuracy in vowel perception studies when compared to steady-state vowels (c.f. Nearey and Assmann 1986, Hillenbrand et al. 1995, Nearey 1999, Zahorian and Jagharghi 1993). Listeners likely use all the information present in the signal to disambiguate vowel identity, especially when the vowel space is crowded and there is a lot of overlap between categories, as shown in Figure 30 by Peterson and Barney (1952) for American English below:

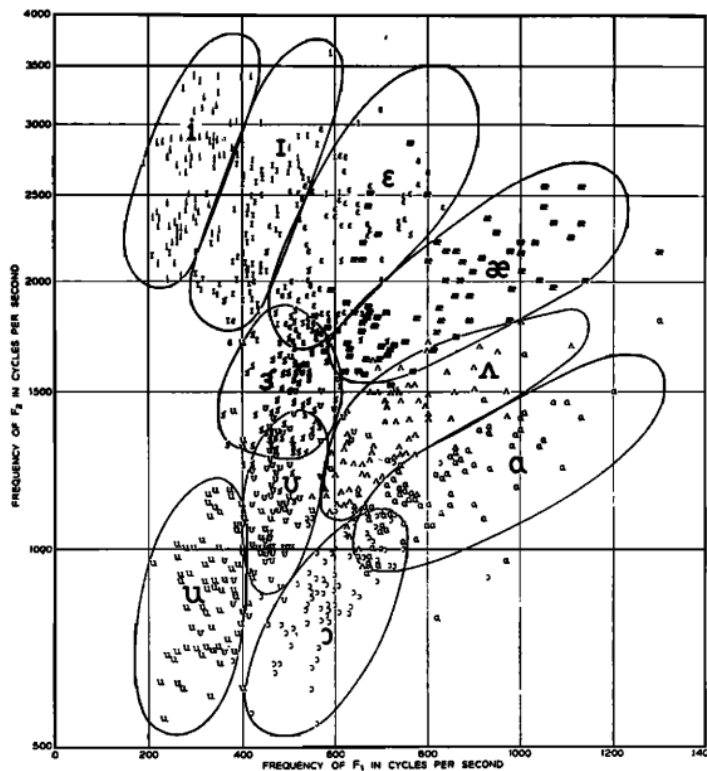


FIG. 8. Frequency of second formant *versus* frequency of first formant for ten vowels by 76 speakers.

Fig. 30: Vowel chart of American English vowels as produced by 76 speakers (Peterson and Barney (1952:182)

There is substantial overlap between vowel categories on the F1/F2 plane. However, if vowels move in different directions and have their own unique pattern of spectral movement, this vowel inherent spectral change (VISC) pattern of movement is likely information listeners utilize to disambiguate. Hillenbrand et al. (1995) have measured F1 and F2 at 20% of the vowel and 80% of the vowel for speakers from the Upper Midwest and have shown spectral change patterns from onset to offset for almost all vowels, with the exception of /i/ and /u/. This is shown in Figure 31 below. While VISC has been shown to provide important information in the disambiguation of vowels

to listeners (c.f. Hillenbrand et al. 1995, Jenkins et al. 1994, Parker and Diehl 1984, Strange 1989, Verbrugge and Rakerd 1986, Simpson, Kohler, Rettstadt 1997, Morrison and Assmann 2013), German has been categorized as a language that shows little VISIC and instead rely more on duration when disambiguating long-short vowel pairs (c.f. Sendlmeier 1981, Strange and Bohn 1998, Morrison and Assmann 2012).

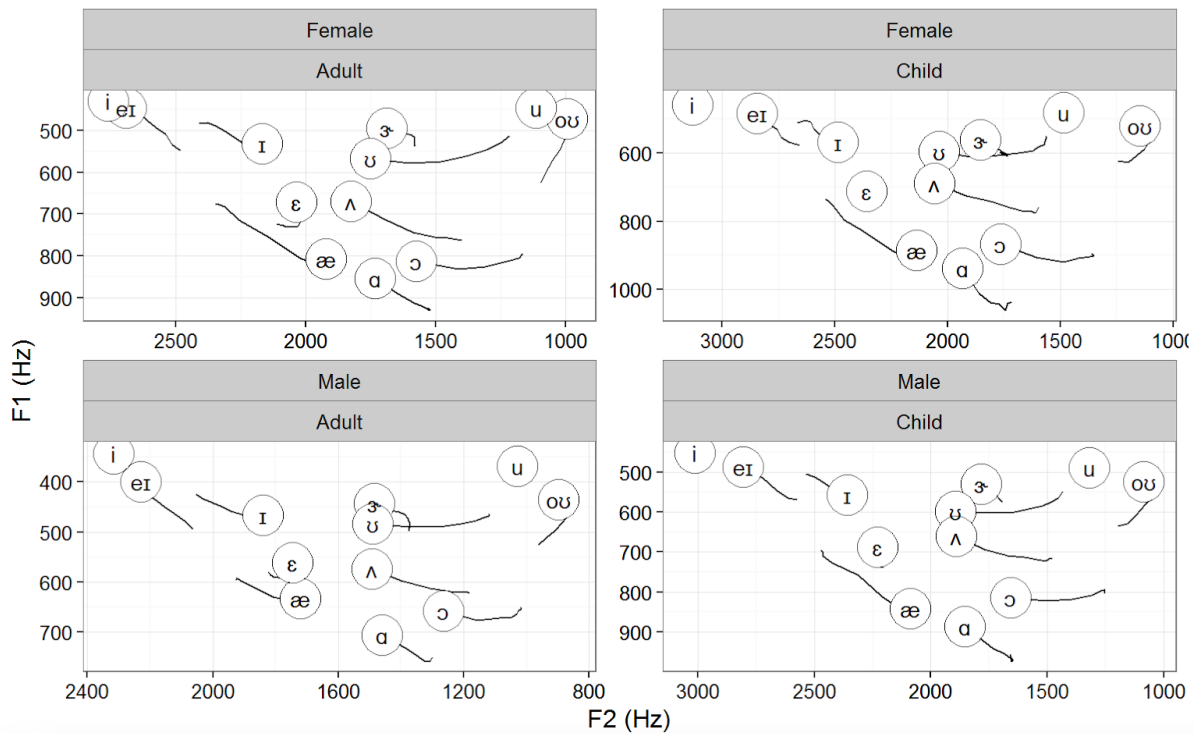


Fig. 31: Vowel trajectories of H95 data plotted by Matt Winn (2016)¹⁷

Hillenbrand et al. (1995) also used a quadratic discriminant analysis to classify 12 vowel types, using fundamental frequency (F0), F1, F2, F3, and VISIC features (20, 50 and 80% of vowel duration) as classifiers. The model showed a high degree of accuracy in identifying the vowel tokens when VISIC information and duration

¹⁷ See: http://www.mattwinn.com/tools/HB95_2.html

information were included in the model. Following up, Hillenbrand and Houde (2003) trained different models on the H95 vowels using either single slice information (from 15, 30, 45, 60, or 75% of the vowel), two slice information (from 15 and 75% of the vowel), or three slice information (from 15, 30 and 75% of the vowel). While single slice information classification accuracy only reached between 75.5 and 80.4%, accuracy improved substantially with two slice information to 90.6%. Three slice information increased classification accuracy only a little to 91.6%. This result shows that the information in the onset and offset of the vowels is critical for identification. VISC information has in fact been shown to be as important in vowel identification as information from the midpoint of the vowel is: Jenkins et al. (1983) have shown used silent-center vowels, showing that classification accuracy was almost as high for those (92.4%) as for the full information vowels (93.1%) for American English. Therefore, the importance of VISC cannot be ignored for German, and information from the onset and offset of the vowels is likely used in vowel perception. The fourth experiment in this dissertation will investigate the degree to which VISC is used in vowel perception by native German-speaking listeners.

4.2 Experiment four

Traditionally German is thought to rely less on dynamic and more on static cues as it is said to not diphthongize monophthongs in comparison to American English and are pure in quality (Strange and Bohn 1998, Strange et al. 2004). If this is the case, flat

formant identifications should not differ from dynamic identifications. Traditional theories of vowel perception in German assume that the main information used to identify vowel quality is contained in static target formant frequency values for F1 and F2. (Strange and Bohn 1998, Strange et al. 2004, Schwartz 2021) However, for English, many studies have shown that time-varying information is used to disambiguate vowels (see Strange 1987, Nearey and Hillenbrand 1999).

Hillenbrand et al. (1995) trained a quadratic discriminant analysis to look at vowel identification rates for steady-state formant values and identification rates for 20% and 80% of the vowel. The steady-state accuracy was at 71% whereas the onset-offset identification rates were at 91%. This falls in line with Nearey's perception model of pattern recognition and is evidence of the importance of VISCs in vowel perception.

To test the importance of VISCs in German, native German listeners were presented with silent center and silent onset-offset vowel tokens. If German relies on static target formants, the silent onset-offset condition should have lower correct identification rates than the silent center vowel instances.

4.2.1 Methods and Materials

4.2.1.1 Listeners

61 native speakers of German were recruited via Linguist List for this study (m = 21, f = 36, non-binary = 4, mean age = 30.7, age σ = 13.6) and completed the experiment on

Qualtrics. All participants reported German as their native language and all but eleven reported being fluent in one or more other languages. None reported any problems with the experiment platform. A table with detailed demographic information can be found in the appendix.

4.2.1.2 Stimuli

4.2.1.2.1 Production

To investigate whether German vowels show formant movement, acoustic measures from experiment one were taken and plotted in R. The results show clear formant movement for all vowels. Figure 32 shows the pattern of formant movement for F1 and F2 at the vowel onset (measured at 20%), the midpoint (measured at 50%), and the offset (measured at 80%).¹⁸ This is in contrast to earlier literature claiming that there is little formant movement in German vowel productions (Strange and Bohn 1998, Strange et al. 2004). While all vowels show formant movement, some vowels show more movement than others. This is in line with previous research, showing that /i:/, for example, showed relatively little movement, while /u:/ showed more movement (Brandt et al. 2018).

Additionally, the role of duration could be important as experiments two and three have shown that German listeners rely on quantity as a primary cue. Therefore, it is

¹⁸ While the recorded tokens contained a stop-vowel-stop sequence, and therefore formant transitions, these are unlikely to affect perception since the stops were kept the same across all vowels.

assumed that there will be different perception patterns for long vowels and short vowels.

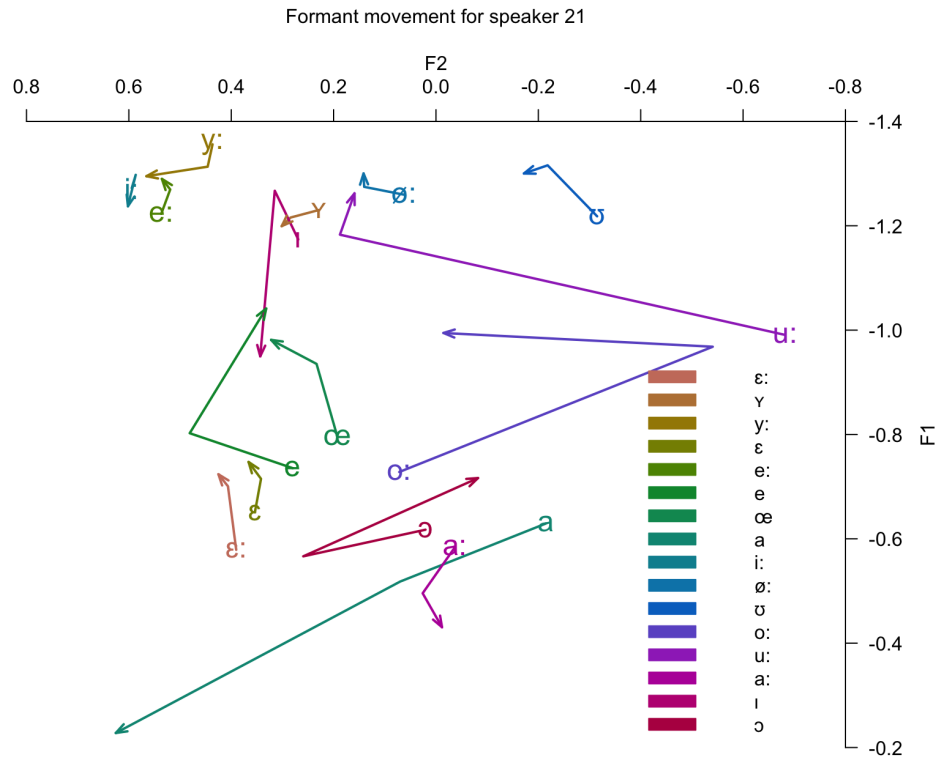


Fig. 32: Formant movement for German normalized F1 & F2 values from production data from experiment one from one female speaker

4.2.1.2.2 Resynthesized Productions

The stimuli for experiment four were re-synthesized using the productions of one male speaker (speaker 19) from experiment 1. Resynthesis was done in Python 3 using the pysptk library¹⁹.

¹⁹ See <https://github.com/r9y9/pysptk>

One male speaker was picked based on recording quality for resynthesis, from these tokens the generalized mel cepstrums were calculated and converted to MGLSADF (Mel-log spectrum approximation digital filter) coefficients, from which speech was then resynthesized.

Resynthesized speech was used because using mel cepstrums allows for describing the “large” structure of the spectrum, focusing on the spectral envelope and relevant formant information, excluding any noise that is not related to the formant structure. This will be described in more detail in the following section but is illustrated below in Figure 33. The coefficients provide information about the formants, therefore ignoring fine spectral structures, specifically filtering out the source information. The cepstrum is the sum of the vocal tract frequency response and the glottal pulse:

$$X(t) = E(t) + H(t)$$

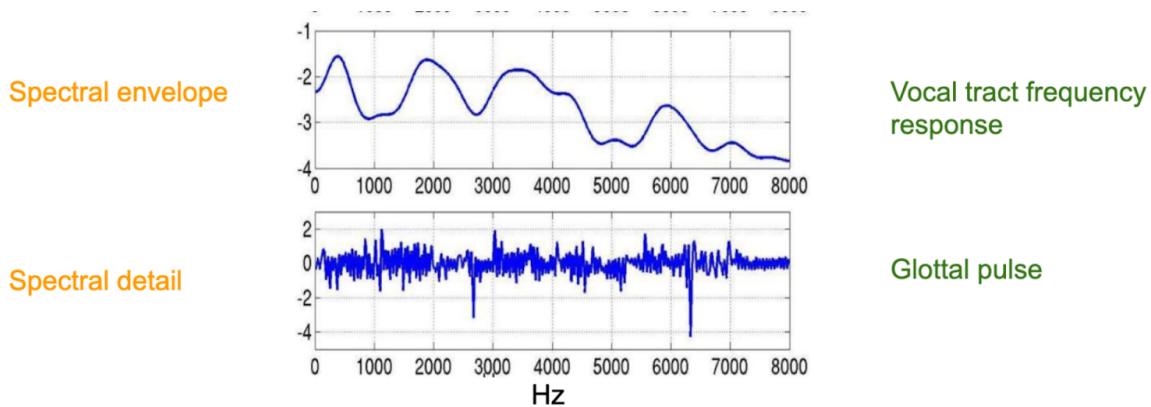


Fig. 33: Understanding the cepstrum (Velardo 2020²⁰)

²⁰ Valerio Velardo: Mel-Frequency Cepstral Coefficients Explained Easily. <https://github.com/musikalkemist/AudioSignalProcessingForML/blob/master/19-%20MFCCs%20Explained%20Easily/Mel-Frequency%20Cepstral%20Coefficients%20Explained%20Easily.pdf>, last accessed 07/15/2021.

Using a low pass filter allows to remove the glottal pulse and the resulting resynthesized speech only contains the cepstral coefficients connected to the spectral envelope and therefore formant information. Using a Mel Filterbank transforms the linear representation to a Mel representation, to which a discrete cosine transform is applied, which is a simplified version of a Fourier transform. The reason for this is to get real-valued coefficients instead of the complex coefficients from an inverse Fourier transform, effectively making it simpler to handle. Another advantage is that it allows for decorrelation of energy in different Mel bands which efficiently reduces the number of dimensions to represent the spectrum. The first 12-13 coefficients are used to preserve the most relevant information: formant information from the spectral envelope. Figure 34 shows the spectral envelope for a mel-generalized cepstrum. Looking at Figures 35 and 36 we can see that information about the glottal pulse, or the fast-changing information in the speech signal, is effectively excluded in the resynthesized waveform, and left is the relevant formant information.

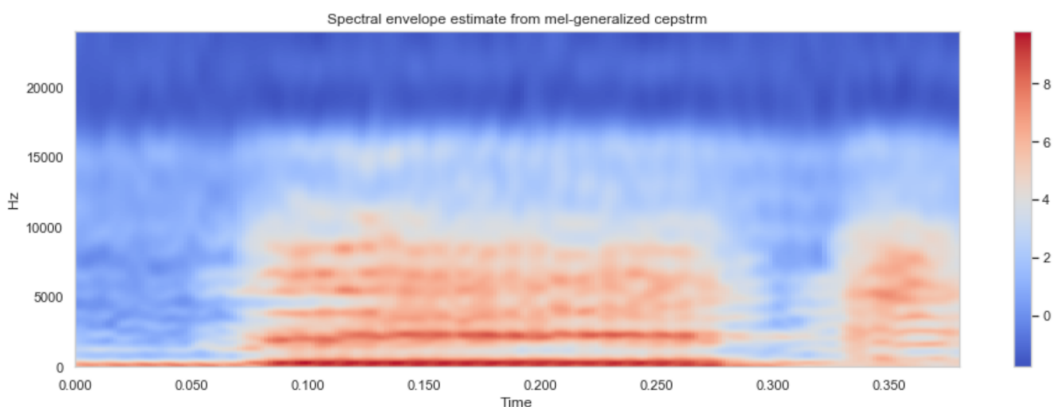


Fig. 34: Spectral envelope from mel-generalized cepstrum for <bäht>

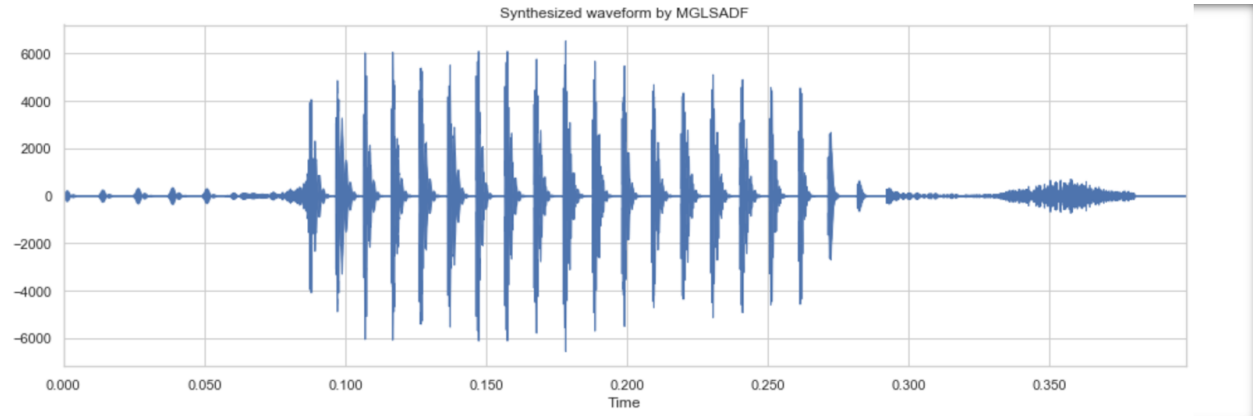


Fig. 35: Resynthesized waveform for <bäht>

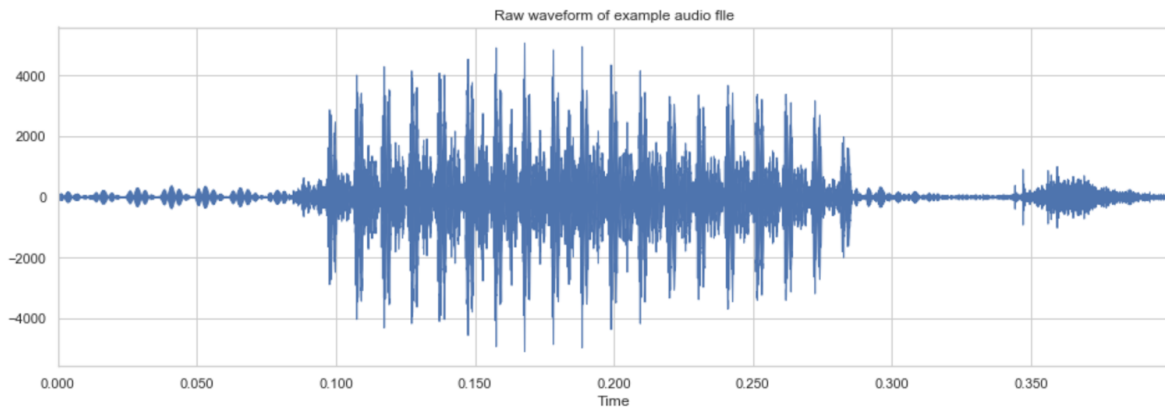


Fig. 36: Raw waveform for <bäht>

All tokens were amplitude normalized to 70dB with a script from Cohn²¹ and had a sampling frequency of 48000 Hz. The tokens were manipulated to contain either silent-center or silent-onset-offset vowels by setting either the middle portion of the vowel or the onset and offset to zero, as shown in the Figures 37 and 38 below.

²¹ See: https://github.com/michellecohn/praat-scripts/blob/master/adjust_mean_intensity_db/, last accessed 04/13/2022.

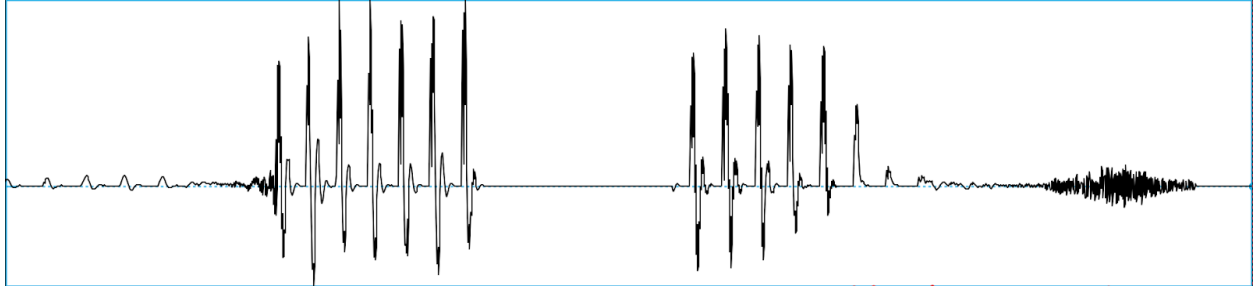


Fig. 37: Silent center token for /bɛ:t/

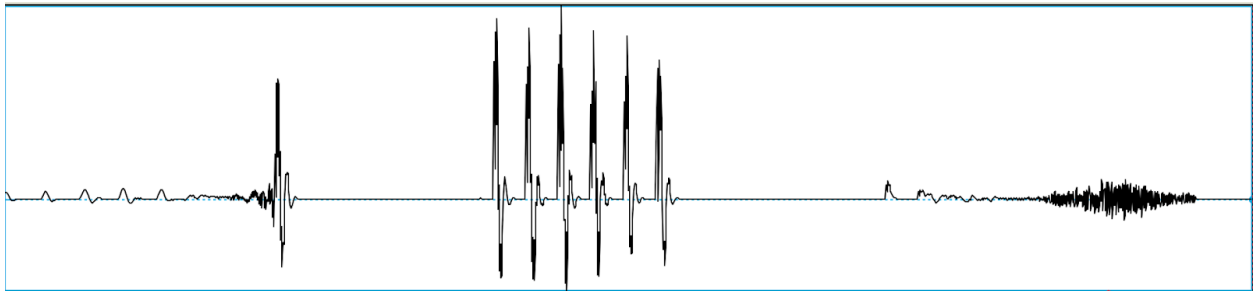
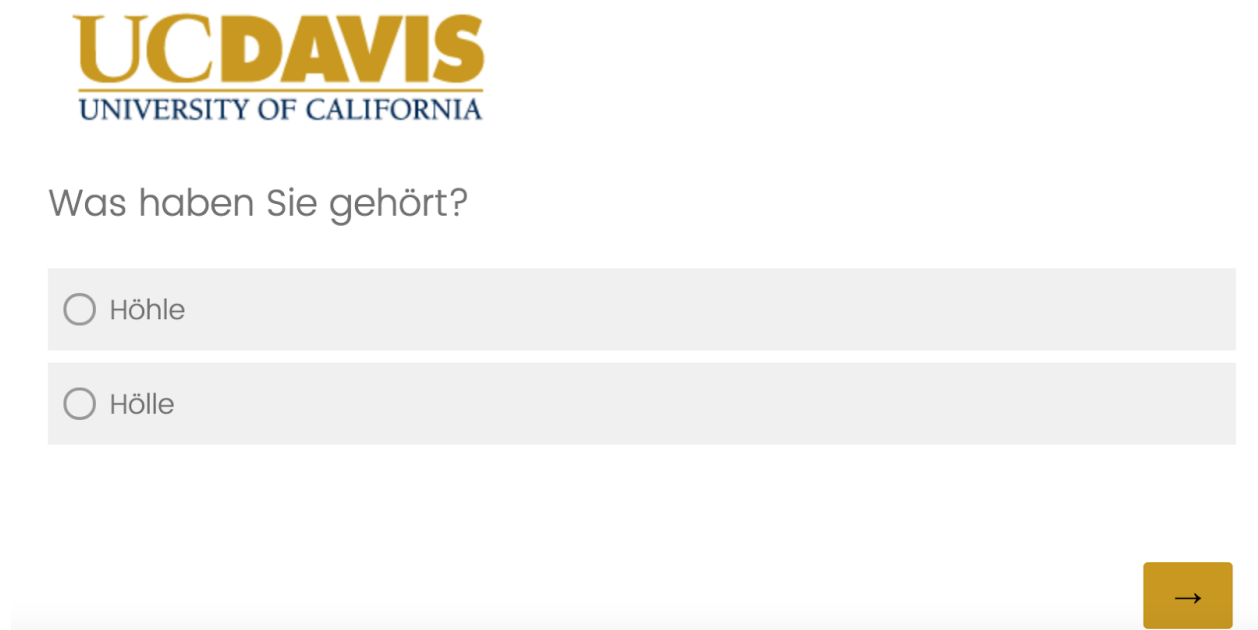


Fig. 38: Silent onset-offset token for /bɛ:t/

4.2.2 Procedure

The experiment was conducted fully online, using the Qualtrics survey platform. Participants were instructed to sit in a quiet room and use a computer to complete the experiment. Before starting the trials, participants were able to play a test sentence and asked to set the volume to a comfortable level to ensure that their sound output worked and the sound was loud enough. The stimuli were presented in randomized order and each token was played once. In total participants responded to 32 tokens (16 vowels X 2 manipulation steps). Participants were presented with a forced-choice task and asked to identify whether a vowel was long or short (e.g. *Baht* vs *Batt*). They were presented

with German word options on the screen that corresponded to the non-words from experiment one. Listeners could not progress to the next trial if they had not chosen a response. Figure 39 shows an example screen.



UC DAVIS
UNIVERSITY OF CALIFORNIA

Was haben Sie gehört?

Höhle

Hölle

→

Fig. 39: Trial screen for /ø:/-/œ/ with minimal pair Höhle - Hölle (cave - hell)

After the perception task participants were asked to fill out a demographic questionnaire.

If formant movement does not matter and the formant values close to the midpoint of the vowel are the main cue in identifying whether a vowel is short or long, the correct identification rates for the silent center tokens should be slightly worse than those for the silent onset offset tokens. However, if VISC is used in German, the correct

identification rates for the silent onset offset tokens should be slightly worse than the silent middle tokens.

4.2.3 Analysis

Responses were coded for accuracy (correct original vowel length chosen = 1, wrong original vowel length chosen = 0). To test the overall importance of VISIC, the data were analyzed using a generalized mixed-effects logistic regression (*lme4* R package; Bates et al., 2015). Main effects included condition (silent middle, silent onset/offset), and original length (long = 1, short = 0), and their interaction. Random effects included by-Listener random intercepts and by-Vowel random intercepts (*lmer* syntax: Accuracy ~ Condition * orig_length + (1 | vowel) + (1 | ID)). Because there were some vowel specific patterns in experiments two and three, vowel specific models were run to investigate the importance of formant movement for each long and short vowel individually. The data were analyzed using Bayesian binary logistic regression models in R with the *brms* package to avoid convergence issues. Main effects included condition (silent middle, silent onset/offset), and original length (long = 1, short = 0), and their interaction. Random effects included by-Listener random intercepts. The priors used were a student's t-distribution ($\nu = 3$, $\mu = 0$, $\sigma = 3$) for the regression coefficients, and a student's t-distribution ($\nu = 3$, $\mu = 0$, $\sigma = 2.5$) for standard deviations of random effects. All models converged (Rhat = 1.0). (*brm* syntax: Accuracy ~ Condition * orig_length + (1 | ID), family = bernoulli(logit))

4.2.4 Results

The output of the logistic regression model is provided in Table 26. There was a significant interaction between condition and original vowel length. The significant interaction was more closely examined with Tukey's HSD pairwise comparisons within the model using the *emmeans()* function in the *emmeans* R package (Lenth et al., 2021). This revealed that listeners had significantly lower correct identification accuracy for long vowels in the silent onset/offset condition than for short vowels ($p = 0.007$) and significantly lower correct identification accuracy for long vowels in the silent onset/offset condition than for long vowels in the silent middle condition ($p < 0.0001$).

No other effects or interactions were observed. As shown in Figure 40, listeners were highly accurate in choosing whether a vowel was long or short, regardless of condition.

Table 26: Regression output for general VISC model

	<i>Est</i>	<i>Std. Err.</i>	<i>z</i>	<i>p</i>
(Intercept)	2.74	0.34	8.05	<0.001
Condition	0.40	0.26	1.54	0.12
OriginalLength	-0.18	0.46	0.39	0.69
Condition:OriginalLength	-1.31	0.33	-3.96	<0.001

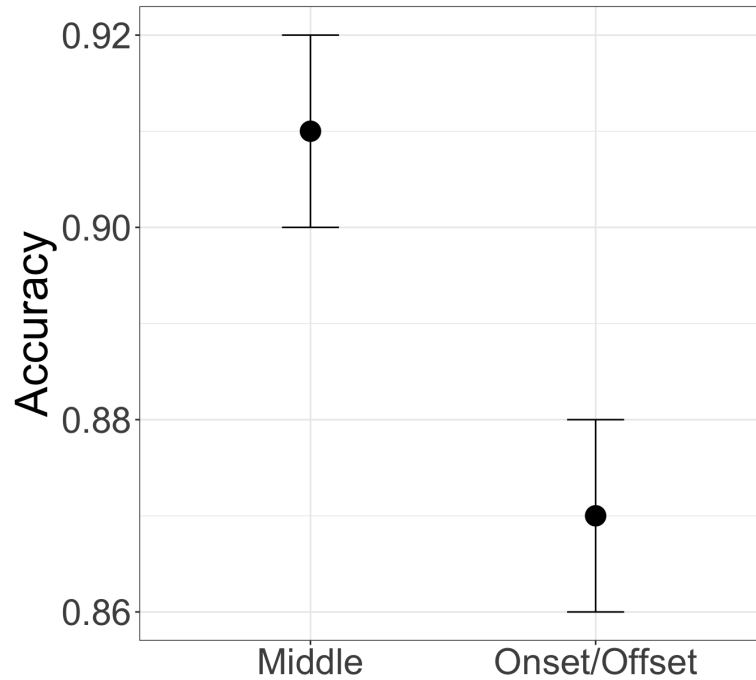


Fig. 40: Listener accuracy by Condition

However, when looking at original vowel length, there was a significant interaction. Figure 41 shows that there was a vowel specific effect with listeners being less accurate in identifying long vowels in the silent onset/offset condition.

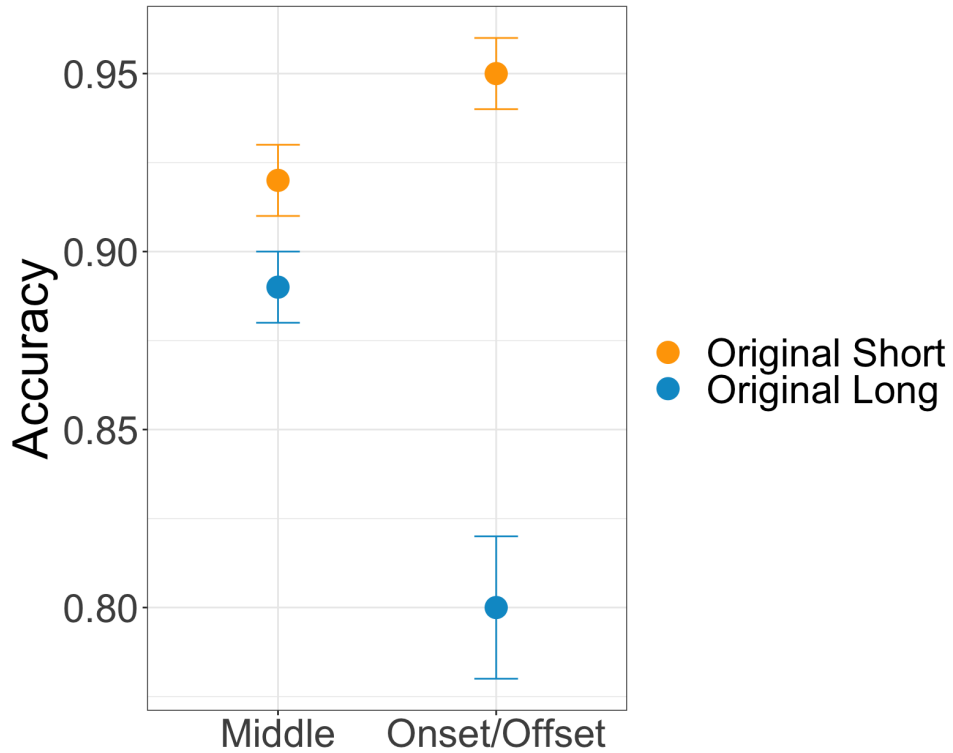


Fig. 41: Listener Accuracy by Condition and Original Length

No significant effects were observed in the the vowel specific models, except for /ʊ/, where the effect of Condition was 1.71 (95% credible interval [.10, 3.64]). Table 27 shows all effects.

Table 27: Brm Output for German /ʊ/

	Estimate	Est.Error	l-95% CI	u-95% CI
Intercept	3.05	1.08	1.52	5.77
ConditionOO	1.71	0.90	0.10	3.64
OrigLength	0.15	5.49	-9.57	9.63
Condition:OrigLength	-0.04	4.82	-9.05	9.33

Recall that in experiment two there were no observed main effects for /u:/-/ʊ/ and in experiment three the effect was not in the expected direction with listeners selecting token as long *less* frequently as continua approached originally long FFS regardless of original duration. Recall also that high back rounded vowels are unstable in the production of FFS and often move into the space of the high front rounded vowels (Hoole and Kühnert 1995, Harrington et al. 2011), and that listeners were insensitive to spectral changes in the /u:/-/ʊ/ space at midpoint (Tomaschek et al. 2015). It then seems that /ʊ/ is a peculiar case and while it might be unstable in the production of FFS, listeners could latch on to the formant trajectories instead of the location in the F1/F2 plane, using dynamic rather than static information.

4.3 Discussion

The results from experiment four provide evidence for listeners using VISC information only in the perception of long vowels in German.²² While overall identification accuracy was still high in both silent center and silent onset/offset conditions, accuracy was overall lower for the silent onset/offset condition, albeit not significantly. Instead, vowel length seems to play a role with accuracy being lower for long vowels in the silent onset/offset condition, as revealed by Tukey's pairwise comparison. As shown in figure 42, short vowels are produced more centrally than long vowels.

²² It is also possible that the consonant specific formant transitions could provide additional information about vowel quality to listeners, however, this was not looked at in this dissertation. Future research could address this question and investigate the importance of consonant to vowel formant transitions.

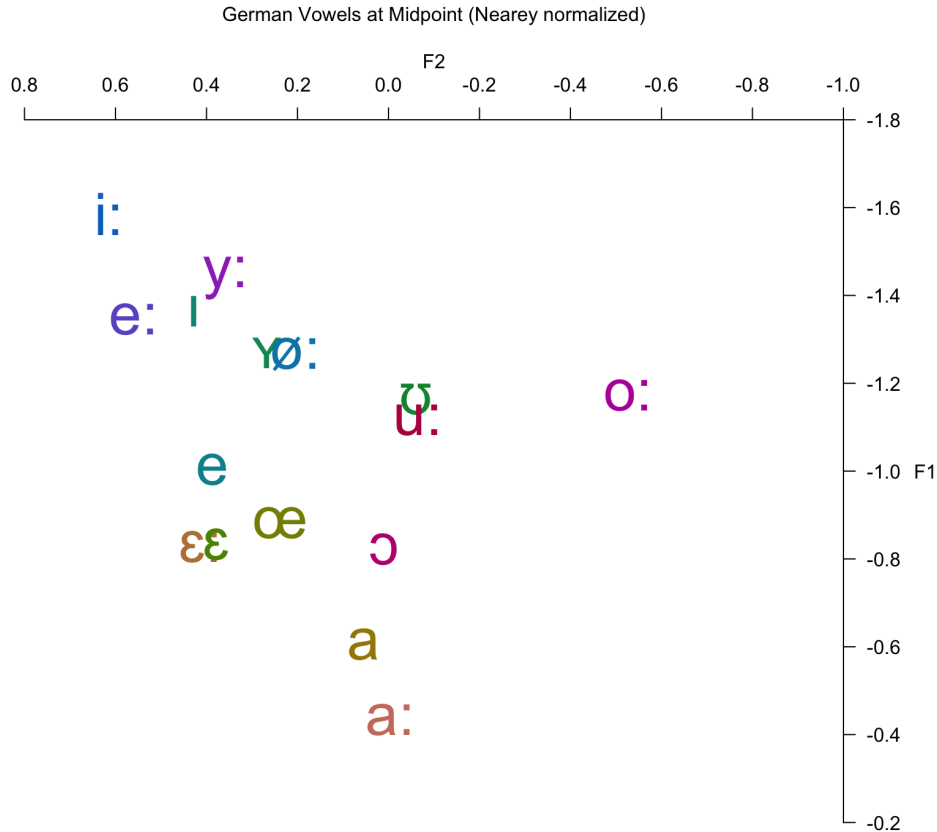


Fig. 24: German vowels at F1/F2 midpoints

Some research has suggested that in the case of short vowels, there is not enough time for the tongue to reach its target and therefore short vowels show a higher degree of undershoot (Diesch et al. 1999, Johnson, Flemming & Wright, 1993) and therefore might show less reliable formant trajectories than their long counterparts. This could translate into VISC information being more heavily exploited in long vowels and therefore the absence of VISC information in long vowels being more detrimental to their correct identification.

Taken together, the results from experiment four provide initial evidence that German listeners show a complex pattern of cue usage in the perception of length in

vowels. It seems that VISC could act as an enhancing cue to further allow listeners to decide whether a vowel was long or short. German listeners' accuracy in long/short identifications dropped significantly for long vowels when VISC information was obscured in the silent onset/offset condition.

Chapter 5

5.1 General Conclusion

This dissertation set out to investigate the cues present in the production and perception of German long and short vowels. In particular, the question of whether spectral differences were present between long and short vowel pairs in production and whether they were salient enough to be used in perception by naive listeners as well as by non-native listeners was explored. In order to test these questions, a production study and multiple perception studies were set up.

The production study showed that all long-short vowel pairs differed in their spectral qualities within a F1/F2 plane in addition to their durations.²³ In order to assess whether these spectral differences were salient enough to be used in perception regardless of the learned perceptual contrasts that are present for L1 German speakers, the German vowels were presented to naive AE listeners in a forced identification and rating task. AE listeners rely on spectral features as cues in their L1 perception (c.f. Joos 1948, Delattre et al. 1952, Peterson 1952, 1961), so the hypothesis for the first perception experiment was that if the spectral features were salient enough, AE listeners would perceive the vowels in a long-short pair as different vowels mapped to their L1. This hypothesis was supported by the data, with AE listeners perceiving all German long-short pairs as different vowels except for German /ɛ:/-/ɛ/ and /u:/-/ʊ/. For the /ɛ:/-/ɛ/ case, this was likely due to the fact that /ɛ:/ is undergoing a merger in process (c.f. Sendlmeier and Seebode 2006, Schoormann et al. 2019, Predeck et al. 2021,

²³ See appendix 3 for all regression outputs.

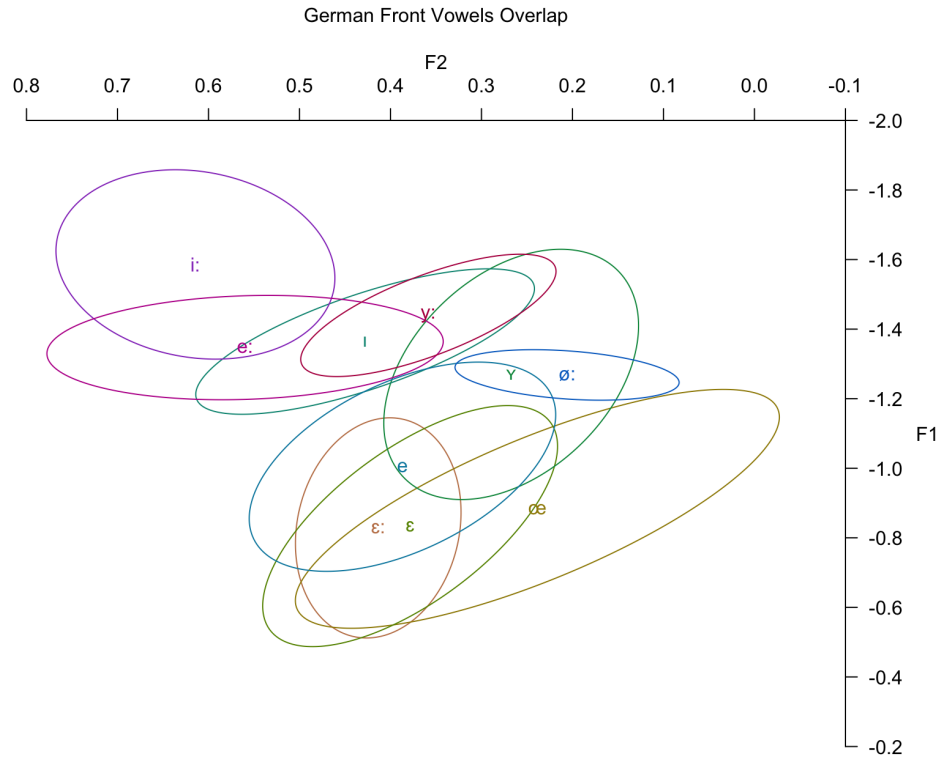
Frank 2021, in prep.) and the region where the speakers were recruited from is the region of Germany where this merger is still ongoing. While a previous study has shown /ɛ:/-/ɛ/ to differ significantly in F1 (Predeck et al. 2021), it differed significantly in F2 in this dissertation.²⁴ This difference might have not been enough for the AE listener to exploit and in this study, and F1 did not differ significantly, which is a cue AE listeners relied on heavily in the identification of the non-native vowels. In the /u:/-/ʊ/ case, a flaw in the experimental design was likely the cause for the very small differences seen in spectral and durational characteristics. The spelling of the short target word *But* in this experiment likely caused listeners to produce this vowel as somewhere between the long *Buht* and the accurate short spelling of *Butt*.

In experiments two and three the usage of quantity and quality was further explored by using five-step duration and spectral continua while keeping the other cue constant. Experiment two used continua that were manipulated for duration but kept the original FFS steady, which resulted in two continua per vowel pair having the same five-step duration differences but either originally long or originally short FFS. The results from experiment two provide evidence for listeners using duration as a primary cue and FFS as a secondary cue. While the proportion of long vowel responses dropped overall as continua were approaching the short duration, listeners were still more likely to select tokens as long if they contained originally long FFS. Experiment three used continua that were manipulated for the first three formant frequencies but kept the original duration steady, which resulted in two continua per vowel pair having the same five-step FFS differences but containing either the originally long or the

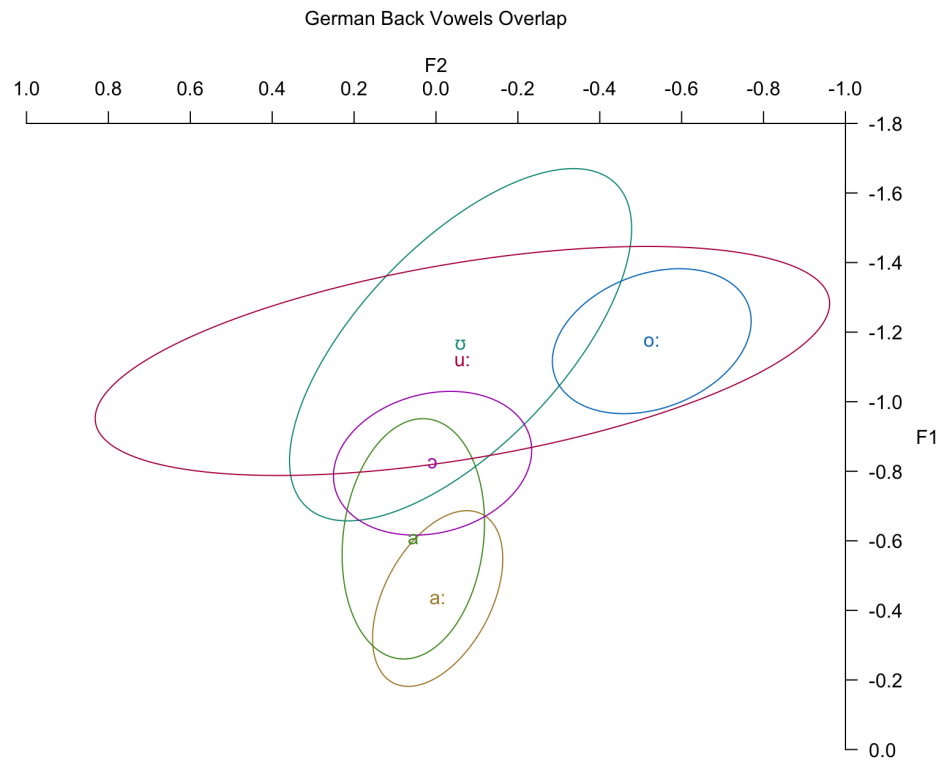
²⁴ These differences are likely due to a merger in progress, in which /ɛ:/ is merging into /e:/. Therefore, acoustic variations in production are expected. For an in-depth discussion of this merger see Frank (2021).

originally short durations. The results from experiment three provide strong evidence for German listeners using duration as a primary cue in the distinction between long and short vowels with listeners selecting the tokens that contained the long durations almost always as long, regardless of FFS manipulation. This shows a strong weighting of quantity. However, there were vowel-specific effects for the front rounded vowel pairs /y:/-/ʏ/ and /ø:/-/œ/, where listeners relied on FFS as well and as FFS approached the originally long values tokens were perceived as long more even for the tokens containing originally short durations.

To investigate the role of spectral features more closely, the fourth experiment was set up to explore the role of VISC in the perception of long and short vowels. There are not many studies investigating the role of VISC in German (Strange and Bohn 1998, Strange et al. 2004), and none to my knowledge that investigate the role of VISC specifically in the distinction between long and short vowels. Exploring the role of VISC in German further is important to give insight into the weight that FFS have in perception. If German listeners use formant information as a secondary cue, they might also rely on dynamic formant information which can provide unique formant trajectories for each vowel, even if there is substantial overlap in F1 and F2 at the midpoints. The production study replicated patterns found by Strange et al. (2007), showing that especially the front vowels, both rounded and unrounded, show substantial overlap in F1 and F2 at midpoints while the back vowels show less overlap. Figure 43 shows a comparison of spectral overlap for front vowels and back vowels using ellipses.



(a)



(b)

Fig. 43: Overlap of German front vowels (a) and German back vowels (b)²⁵

The experiment used silent middle and silent onset/offset tokens to investigate whether German listeners utilized VISC. Results show that listeners were significantly less accurate in the silent onset/offset condition for long vowels, supporting that VISC is used in the disambiguation of long vowels. The conclusion that can be drawn from this is that the spectral features of German vowels are not contained in one single point, and as shown in figure VISCGER all German vowels show formant movement. These patterns aid listeners in the separation of German vowels spectrally in a vowel space that shows a lot of overlap at the midpoint, as shown in figure DENSE. It is not surprising that German listeners rely on multiple cues in order to disambiguate whether a vowel is long or short with a relatively dense vowel space. However, short vowels in German are on average only 0.28 seconds long. Therefore, the tongue might not reach its target and short vowels are likely to show a higher degree of undershoot (Diesch et al. 1999, Johnson, Flemming & Wright, 1993). Formant trajectories for short vowels could be uninformative at best and ambiguous at worst in comparison to long vowels.

Recall that the main goal of this dissertation was to establish whether German listeners rely on quantity or quality as a primary cue when disambiguation whether a vowel is long or short within a vowel pair and what the role of secondary cues is. The quantity-quality debate states that either quality *or* quantity is distinctive while the other one is redundant (c.f. Riad 1995, Wiese 1996, Lahiri and Drescher 1999, Vennemann

²⁵ Note that values for /u:/-/ʊ/ are not representative due to a flaw in the experimental design, as mentioned in chapter two.

2000, Mangold 1990, Delattre 1969, Vernon 1976, Maack 1951, 1954, Jessen 1993, Weiss 1977) and previous studies have shown evidence for either quality being used in the perception of long/short vowel pairs (Mangold 1990, Delattre 1969, Vernon 1976), *or* quantity (Maack 1951, 1954). The data from these experiments instead show that while quantity is used as a primary cue, quality is used as a secondary cue. Additionally, there was also evidence suggesting vowel-specific patterns with the front rounded vowels providing strong evidence for the use of quality in experiment three. This means that neither cue is used alone, and instead German listeners seem to employ vowel specific cue usage patterns. This complex pattern of cue usage is also supported by the data from experiment four which shows that German listeners use VISC in the identification of long vowels, but not in the identification of short vowels.

Thus, the final conclusion of this dissertation is that while German listeners use quantity primarily for all but front-rounded vowels, quality is used as a secondary, enhancing cue. Additionally, German listeners rely on VISC information in the perception of long vowels, further supporting the secondary use of spectral information. Duration, then, is not the sole cue used in the perception of German long/short vowel pairs. This dissertation provides clear evidence for the use of vowel quality being used as a secondary cue.

References

- Adank, P.M. (1972). Vowel Normalization: a perceptual-acoustic study of Dutch vowels.
- Adank, P.M., Smits, R., van Hout, R. (2004). A comparison of vowel normalization procedures for language variation research.
- Ainsworth, W. (1975). Intrinsic and extrinsic factors in vowel judgments. In G. Fant & M. Tatham (Eds.), *Auditory analysis and perception of speech*. London: Academic Press, 103-113.
- Alonso, J. G., Rothman, J., Berndt, D., Castro, T., & Westergaard, M. (2017). Broad scope and narrow focus: On the contemporary linguistic and psycholinguistic study of third language acquisition. *International Journal of Bilingualism*, 21(6), 639-650.
- Appelbaum, I. (1996). The lack of invariance problem and the goal of speech perception. In *Proceedings of the Fourth International Conference on Spoken Language Processing. ICSLP'96 (Vol. 3, pp. 1541-1544)*. IEEE.
- Assmann, P. F., Nearey, T. M. (2007). "Relationship between fundamental and formant frequencies in voice preference," *J. Acoust. Soc. Am.* 122, EL35–EL43.
- Auer, P., Schmidt, J. E., & Lameli, A. (Eds.). (2010). *Language and space: An international handbook of linguistic variation (Vol. 1)*. Walter de Gruyter.
- Baesecke, G. (1930). *Der Deutsche Abrogans und Die Herkunft Des Deutschen Schrifttums*.
- Barbour, S., Stevenson, P.P. (1990). *Variation in German: A critical approach to German sociolinguistics*. Cambridge: Cambridge University Press.
- Barca, L. (2021). Toward a speech-motor account of the effect of age of pacifier withdrawal. *Journal of Communication Disorders*, 90, 106085.
- Barreda, Santiago. "Vowel normalization as perceptual constancy." *Language* 96, no. 2 (2020): 224-254.
- Barreda, S. 2015 phonTools: Functions for phonetics in R. R package version 0.2-2.1.
- Beddor, P. S., & Strange, W. (1982). Cross-language study of perception of the oral-nasal distinction. *Journal of the Acoustical Society of America*, 71(6), 1551e1561.

- Behne, D. M., Czigler, P. E, Sullivan, K. P. H. (1996). Acoustic characteristics of perceived quantity and quality in Swedish vowels, *Speech Science and Technology '96*, Adelaide, Australia; 49-54, 1996.
- Behne, D., Czigler, P., & Sullivan, K. P. (1997). Swedish quantity and quality: a traditional issue revisited. *Reports from the Department of Phonetics, Umeå University*, 4, 81-83.
- Benediktsson, J.; Swain, P.; Ersoy, E. (1990). Neural Network Approaches Versus Statistical Methods in Classification of Multisource Remote Sensing Data. *IEEE Transactions on Geoscience and Remote Sensing* 28.4.
- Bennett, D. (1968). Spectral form and duration as cues in the recognition of English and German vowels. *Language and Speech*, 11, 65-81.
- Benson, H. (2001). *Principles of Vibration*. Second Edition, Oxford University Press.
- Berkes, E., Flynn, S. (2012). "Further evidence in support." *Third language acquisition in adulthood* 46 (2012): 143.
- Berns, M. (1988). The cultural and linguistic context of English in West Germany. *World Englishes*, 7(1), 37-49.
- Birdsong, D. (1992). Ultimate attainment in second language acquisition. *Language*, 706-755.
- Bleeck, S., O'Meara, N. (2014). The perception of size and shape of resonant objects. *Journal of Hearing Science*, 1-12.
- Bloomfield, L. (1933). *Language*. New York: Holt.
- Blumstein, S. E. (2021). Features in Speech Perception and Lexical Access. In: Pardo, J. S., & Nygaard, L. C. (2021). *The Handbook of Speech Perception*.
- Bohn, O. S., Flege, J. E. (2005). "Acoustic and perceptual similarity of North German and American English vowels." *The Journal of the Acoustical Society of America* 115.4 (2004): 1791-1807.
- Bohn, O.S., Flege, J.E. (1990). Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics* 11:303-28
- Bohn, O.S., Polka, L. (2001). Target spectral, dynamic spectral, and duration cues in infant perception of German vowels. *The Journal of the Acoustical Society of America*, 110(1), 504-515.

- Bohn, O.S.; Polka, L. (2003). Asymmetries in vowel perception. *Speech Communication* 41.221-231.
- Bongaerts, T., van Summeren, C., Planken, B., Schils, E. (1997). Age and ultimate attainment in the pronunciation of a foreign language. *Studies in Second Language Acquisition*, 19, 447–465.
- Box, G. E., Tiao, G. C. (2011). *Bayesian inference in statistical analysis* (Vol. 40). John Wiley & Sons.
- Bradlow, A.R., Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition* Volume 106, Issue 2, February 2008, Pages 707-729
- Brandt, E., Zimmerer, F., Andreeva, B., Möbius, B. (2018). Impact of prosodic structure and information density on dynamic formant trajectories in German. In 9th International Conference on Speech Prosody (pp. 119-123).
- Brokoff, J. (2007). Poesie und Grammatik: der Anteil der Sprachgesellschaften an der Entwicklung der deutschen Literatur- und Poesiesprache in der ersten Hälfte des 18. Jahrhunderts. In: Jean-Marie Valentin (Hg.): Akten des XI. Internationalen Germanistenkongresses Paris 2005 »Germanistik im Konflikt der Kulturen«. Bd. 5: Kulturwissenschaft vs. Philologie? Bern u. a., 229–238.
- Brown, A. (1995). Minimal pairs: minimal importance?. *ELT Journal*, 49(2), 169-175.
- Bruderer, A. G., Danielson, D. K., Kandhadai, P., & Werker, J. F. (2015). Sensorimotor influences on speech perception in infancy. *Proceedings of the National Academy of Sciences*, 112(44), 13531-13536.
- Bruner, J. S. (1973). Pacifier produced visual buffering in human infants. *Developmental Psychobiology: The Journal of the Intern*
- Bürkner, P. C. (2016). “brms: An R package for Bayesian multilevel models using Stan”. In: *Journal of Statistical Software* 80.1, pp. 1– 28.
- Calvo Fernández, M. (2018) *Vowel change in English and German: a comparative analysis*.
- Campus, T. V. L. (2012, December 17). PHY230 - The Sound System of German [Video]. YouTube. <https://www.youtube.com/watch?v=uc-mtGPD3-U&feature=youtu.be>
- Cardoso, A., Hall-Lew, L., Kementchedjheva, Y., Purse, R. (2017). Between California and the Pacific Northwest: The front lax vowels in San Francisco english. *American Speech: A Quarterly of Linguistic Usage*, 91(Supplement (101) 1), 33-54.

- Carello, C., Wagman, J. B., Turvey, M. T. (2005). Acoustic specification of object properties. *Moving image theory: Ecological considerations*, 79-104.
- Carlson, R., Fant, G., Granström, B. (1975). Two-formant models, pitch and vowel perception. In *Auditory analysis and perception of speech*, pp. 55-82.
- Chambers, W. W., Wilkie, J. R. (2014). *A Short History of the German Language (RLE Linguistics E: Indo-European Linguistics)*. Routledge.
- Charlton, B. D., Ellis, W. A., Larkin, R., Fitch, W. T. (2012). Perception of size-related formant information in male koalas (*Phascolarctos cinereus*). *Animal cognition*, 15(5), 999-1006.
- Charlton, B. D., Reby, D., McComb, K. (2007). Female perception of size-related formant shifts in red deer, *Cervus elaphus*. *Animal Behaviour*, 74(4), 707-714.
- Charlton, B. D., Zhihe, Z., Snyder, R. J. (2009). The information content of giant panda, *Ailuropoda melanoleuca*, bleats: acoustic cues to sex, age and size. *Animal Behaviour*, 78(4), 893-898.
- Chomsky, N. (1959). Review of B.F. Skinner, *Verbal behavior*. *Language*, 1959, 35, 26-58.
- Chomsky, N. M. Halle (1968). *The Sound Pattern of English*. Harper and Row: New York
- Clarke, C.M., Garrett, M.F. (2004). Rapid adaptation to foreign-accented English. *The Journal of the Acoustical Society of America*, 116(6), 3647–3658.
- Clay, G. (2008). *1000 Jahre deutsche Literatur*. Hackett Publishing.
- Clopperr, C.G., Tamati, T.N. (2010). Lexical recognition memory across dialects. *The Journal of the Acoustical Society of America* 127, 1956
- Clyne, M. (1995). *The German language in changing Europe*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511620805>.
- Cobb, K., & Simonet, M. (2015). Adult Second Language Learning of Spanish Vowels. *Hispania*, 98(1), 47–60. <http://www.jstor.org/stable/24368851>
- Cook, V. (2017). *Second Language Acquisition: One Person with Two Languages*. *The Handbook of Linguistics*, 557-581.
- Cooper, F. S., Liberman, A. M., Borst, J. M. (1951). The interconversion of audible and visible patterns as a basis for research in the perception of speech. *Proceedings*

- of the National Academy of Sciences of the United States of America, 37(5), 318– 325.
- Cutler, A. (1995). Spoken word recognition and production. In J. L. Miller & P. D. Eimas (Eds.), *Speech, language, and communication* (pp. 97-137). San Diego: Academic Press.
- d'Alquen, R. (1979). Acoustic phonetics and vowel quantity in the history of German. *Zeitschrift für Dialektologie und Linguistik*, 187-204.
- de la Fuente, A. A., Lacroix, H. (2015). Multilingual learners and foreign language acquisition: Insights into the effects of prior linguistic knowledge. *Language Learning in Higher Education*, 5(1), 45-57.
- Delack, J. B. (1972). Überlänge in German vowels revisited.
- Delahunty, G. P., Garvey, J. J. (2004). *Phonetics and phonology*. Colorado: Colorado State University.
- Delattre PC, Liberman AM, Cooper FS. (1951). 1951 Voyelles synthétiques à deux formants et voyelles cardinales. *Le Maître Phon.* 96:30– 37
- Delattre PC, Liberman AM, Cooper FS. (1951). 1955 Acoustic loci and transitional cues for consonants. *J. Acoust. Soc. Am.* 27:769–73
- Delattre PC, Liberman AM, Cooper FS. (1964). Formant transitions and loci as acoustic correlates of place of articulation in American fricatives. *Stud. Linguist.* 18:104–21
- Delattre PC, Liberman AM, Cooper FS. (1952). An experimental study of the acoustic determinants of vowel color. *Word* 8:195–210
- Delattre, P. (1969). An acoustic and articulatory study of vowel reduction in four languages. In Malmberg, B. (ed.) 1981. *Pierre Delattre. Studies in comparative phonetics. English, German, Spanish and French*. Heidelberg: Groos. 63-93.
- Delattre, P. C., Liberman, A.M., Cooper, F. S., Gerstman, L. J. (1952). "An experimental study of the acoustic determinants of vowel colour: Observations on one- and two-formant vowels synthesized from spectrographic patterns," *Word* 8, 195-210.
- Dempster, A. P. (1968). A generalization of Bayesian inference. *Journal of the Royal Statistical Society: Series B (Methodological)*, 30(2), 205-232.
- Deng, L., Lenning, M., Mermelstein, P. (1989). Use of vowel duration information in a large vocabulary word recognizer. *Journal of the Acoustic Society of America*, 86, 540-8.

- Diehl, R. L., Lotto, A. J., Holt, L. L. (2004). Speech perception. *Annu. Rev. Psychol.*, 55, 149-179.
- Disner, S.F. (1980). Evaluation of vowel normalization procedures, *J. Acoust. Soc. Am.* 67, 253–261
- Dorman, M. F., Studdert-Kennedy, M. Raphael, L. J. (1977). Stop consonant recognition: Release bursts and formant transitions as functionally equivalent, context-dependent cues. *Perception & Psychophysics*, 22, 109- 122.
- Duden (1990). Duden Aussprachewörterbuch: Wörterbuch der deutschen Standardaussprache
- Dudenredaktion (2015). Duden - Das Aussprachewörterbuch: Betonung und Aussprache von über 132.000 Wörtern und Namen Volume 16 of Duden - Deutsche Sprache in 12 Bänden. Edition 7. Bibliographisches Institut GmbH, 2015.
- Dunn, H. K. (1950). "The Calculations of Vowel Resonances, and an Electrical Vocal Tract," *Journal of the Acoustical Society of America*, 22.740-53
- Durand, J. (2005). Tense/lax, the vowel system of English and phonological theory. *Headhood, elements, specification and contrastivity: Phonological papers in honor of John Anderson*. Philadelphia, PA: John Benjamins, 77-98.
- Elert, C-C. (1964). *Phonologic studies of quantity in Swedish*. (Almqvist & Wiksell: Stockholm).
- Ellis, R. (1989). *Understanding second language acquisition* (Vol. 31). Oxford: Oxford university press.
- Elmiger, D. (2019). *Deutsch undeutlich: eine Begriffsreise durch die vielfältige deutsche Sprache in der Schweiz*.
- Escudero, P. (2000). The perception of English vowel contrasts: Acoustic cue reliance in the development of new contrasts. In *Proceedings of the 4th International Symposium on the Acquisition of Second-Language Speech, New Sounds* (pp. 122-131).
- Escudero, P., Bion, R. A. H. (2007). Modeling vowel normalization and sound perception as sequential processes. In *Proceedings of the 16th international congress of phonetic sciences* (pp. 1413-1416).
- Escudero, P., Benders, T., Lipski, S.C. (2009). Native, non-native and L2 perceptual cue weighting for Dutch vowels: The case of Dutch, German, and Spanish listeners.

- Fant, G., Henningsson, G., Stålhammar, U. (1969). Formant frequencies of Swedish vowels. Dept. for Speech, Music and Hearing Quarterly Progress and Status Report 10.4.26-31.
- Flynn, S., Foley, C., Vinnitskaya, I. (2004). The cumulative-enhancement model for language acquisition: Comparing adults' and children's patterns of development in first, second and third language acquisition of relative clauses. *International Journal of Multilingualism*, 1(1), 3-16.
- Fowler, C. A. (1996). Listeners do hear sounds, not tongues. *J. Acoust. Soc. Am.* 99:1730–41
- Fowler, C. A., Magnuson, J. S. (2012). *Speech perception*.
- Fox, R. A. (1989). Dynamic Information in the Identification and Discrimination of Vowels. *Phonetica* 46.97-116.
- Frank, M. 2021 Merger or near-merger? Acoustic analyses of /e:/ and /ɛ:/ in spoken Standard German, TABU Dag 2021.
- Frank, M. (2021). (in prep.). Phonemzusammenfall im gesprochenen Standarddeutschen? Experimentalphonetische Untersuchungen zu /e:/ und /ɛ:/ [working title; PhD dissertation]. Oldenburg.
- Gabry, J., Mahr, T. (2022). “bayesplot: Plotting for Bayesian Models.” R package version 1.9.0, <https://mc-stan.org/bayesplot/>.
- Garcia, G. D. (2018). *Plotting vowels in R*.
- Giles, M. B., Nagapetyan, T., Ritter, K. (2015). Multilevel Monte Carlo approximation of distribution functions and densities. *SIAM/ASA Journal on Uncertainty Quantification*, 3(1), 267-295.
- Glidden, C., Assmann, P. F. (2004). “Effects of visual gender and frequency shifts on vowel category judgments,” *Acoust. Res. Lett. Online* 5, 132–138.
- Gloning, T., & Young, C. (2004). *A history of the German language through texts*. Routledge.
- Goblirsch, K. (2018). *Gemination, lenition, and vowel lengthening: On the history of quantity in Germanic* (Vol. 157). Cambridge University Press.
- Goddard, C., Wierzbicka, A. (2002). *Meaning and universal grammar: Theory and empirical findings* (Vol. 1). John Benjamins Publishing.

Guion et al 2000

Hadding-Koch, K., Abramson, A.S. (1964). Duration versus spectrum in Swedish vowels: some perceptual experiments. *Studia Linguistica*, 1964:2, 94-107 .

Hagiwara, R. (2009). So what's the deal with formants?
<https://home.cc.umanitoba.ca/~robh/howto.html>

Halle, M. (1977). "Tenseness, vowel shift, and the phonology of back vowels in Modern English," *Linguistic Inquiry* 8: 611–626.

Hanhardt, A. M., Obrecht, D. H., Babcock, W. R., Delack, J. B. (1965). A spectrographic investigation of the structural status of Überlänge in German vowels. *Language and speech*, 8(4), 214-218.

Hardcastle, W. J., Laver, J., & Gibbon, F. E. (Eds.). (2012). *The handbook of phonetic sciences* (Vol. 119). John Wiley & Sons.

Harrington, J., Hoole, P., Kleber, F., Reubold, U. (2011). The physiological, acoustic, and perceptual basis of high back vowel fronting: Evidence from German tense and lax vowels. *Journal of Phonetics*, 39(2), 121-131.

Hawkins, J. A. (2009). German. In *The world's major languages* (pp. 101-124). Routledge.

Hay, J., Warren, P., Drager, K. (2006). Factors influencing speech perception in the context of a merger-in-progress. *Journal of phonetics*, 34(4), 458-484.

Heid, S.; Wesenick, M. B. (1995). Phonetic analysis of vowel segments in the PhonDat base of spoken German.

Heike, G. (1969). *Suprasegmentale Analyse*. Elwert, Marburg.

Heike, G. (1970). Lautdauer als Merkmal der wahrgenommenen Quantität, Qualität und Betonung im Deutschen. *Proc. 6th Int. Congr. Phonet. Sci.*, 1970, pp. 433-437.

Heike, G. (1972). Quantitative und aualitative Differenzen von /a(:)/-Realisationen im Deutschen. *Proc. 7th Int. Congr. Phonet. Sci.*, 1972, pp. 725-729.

Hickey, R. (2018). 'Yes, that's the best': Short front vowel lowering in English today: Young people across the anglophone world are changing their pronunciation of vowels according to a change which started in North America. *English Today*, 34(2), 9-16.

Hickok, G. (2010). The role of mirror neurons in speech perception and action word

semantics. *Language and cognitive processes*, 25(6), 749-776.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical society of America*, 97(5), 3099-3111.

Hillenbrand, J. and Gayvert, R.T. (1993). Vowel classification based on fundamental frequency and formant frequencies. *Journal of Speech and Hearing Research*, 36, 694-700.

Hillenbrand, J. M. (2013). Static and dynamic approaches to vowel perception. In *Vowel inherent spectral change* (pp. 9-30). Springer, Berlin, Heidelberg.

Hillenbrand, J. M., Clark, M., Houde, R. (2000). Some effects of duration on vowel recognition. : *The Journal of the Acoustical Society of America* 108.3013.

Hillenbrand, J., Gayvert, R. T. (1993). Vowel classification based on fundamental frequency and formant frequencies. *Journal of Speech, Language, and Hearing Research*, 36(4), 694-700.

Hillenbrand, J.M., Nearey, T. (1999). Identification of resynthesized /hVd/ utterances: Effects of formant contour. *The Acoustical Society of America*.

Hillenbrand, J.M., Houde, R.A (2003). A narrow band pattern-matching model of vowel perception. *J. Acoust. Soc. Am.* 113, 1044–1055

Hintzman, D. (1986). "Schema abstraction" in a multipletrace memory model. *Psych. Review*, 93, 411-428.

Hoffman, A. (2011). A comparison of native and non-native vowels featuring German as a target language: A statistical analysis using corpora, Tübingen, Germany. <<http://nbn-resolving.de/urn:nbn:de:bsz:21-opus-54494>>.

Holland, C. (2014). *Shifting or Shifted? The state of California vowels*. University of California, Davis.

Hoole, P., Kühnert, B. (1995). Patterns of lingual variability in German vowel production. *Proc. of XIIIth ICPHS, Stockholm*, 2, 442-446.

Horan, G., Langer, N., Watts, S. (2009). *Landmarks in the History of the German Language* (No. 52). Peter Lang.

House, A. S. (1961). On vowel duration in English. *The Journal of the Acoustical Society of America*, 33(9), 1174-1178.

- Ito, M., Tsuchida, J., & Yano, M. (2001). On the effectiveness of whole spectral shape for vowel perception. *Journal of the Acoustical Society of America*, 110, 1141–9.
- Iverson, G. K., Davis, G. W., Salmons, J. C. (1994). Blocking environments in Old High German umlaut. *Folia linguistica historica*, 28(Historica-vol-15-1-2), 131-148.
- Jaensch, C. (2013). Third language acquisition: Where are we now?. *Linguistic Approaches to Bilingualism*, 3(1), 73-93.
- Jakobson, R., Fant, G., Halle, M. (1952). *Preliminaries to speech analysis* (sixth printing 1965). Cambridge, Mass.: MIT Press.
- Jakobson, R., Fant, G., Halle, M. (1952). *Preliminaries to speech analysis*. Cambridge: MIT Press.
- Jakobson, R., Halle, M. (1968). Phonology in relation to phonetics. In Malmberg, B. (ed.) *Manual of Phonetics*, Amsterdam: North Holland. 411-449.
- Jenkins J.J., Strange, W., Edman, T.R. (1983). Identification of vowels in ‘vowelless’ syllables. *Percept Psychophys* 34, 441–450
- Jenkins, J. J., Strange, W., Miranda, S. (1994). ‘Vowel identification in mixed-speaker silent-center syllables,’ *J. Acoust. Soc. Am.* 95, 1030–1043.
- Jessen, M. (1993). Stress conditions on vowel quality and quantity in German. *Working Papers of the Cornell Phonetics Laboratory*, 8, 1-27.
- Johnson, K. (1990). The role of perceived speaker identity in F₀ normalization of vowels. *The Journal of the Acoustical Society of America*, 88(2), 642-654.
- Johnson, K. (1991). Differential effects of speaker and vowel variability on fricative perceptions. *Language and Speech*, 34, 265-279.
- Jones, D. (1917). *An English pronouncing dictionary*. First edition. London: Dent's Modern Languages Series.
- Jones, D. (1918). *An outline of English phonetics*. Leipzig: B.G. Teubner and Cambridge: Heffer and Sons.
- Jongman, A., Fourakis, M., Sereno, J. (1989). The acoustic vowel space of Modern Greek and German. *Language and Speech* 32.3.221-248.
- Joos, M. (1948). *Acoustic Phonetics* [= *Language Monograph* 23, Supplement to *Language* 24: 2].
- Jørgensen, H.P. (1969). *Die gespannten und ungespannten Vokale in der*

- norddeutschen Hochsprache mit einer spezifischen Untersuchung der Struktur ihrer Formantenfrequenzen. *Phonetica*, 19(4), 217-245.
- Jurafsky, D., Martin, J. H. (2019). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*.
- Juscyk, P.W., Bertoncini, J., Bijeljic-Babic, R., Kennedy, L., Mehler, J. (1990). The role of attention in speech perception by young infants
- Karp, R. M., Luby, M., Madras, N. (1986). Monte-Carlo approximation algorithms for enumeration problems. *Journal of algorithms*, 10(3), 429-448.
- Keller, R.E. (1978). *The German Language*. Humanities Press Inc. New Jersey.
- Kennetz, K. (2010). German and German political disunity: an investigation into the cognitive patterns and perceptions of language in post-unified Germany. In "Perceptual Dialectology" (pp. 317-336). De Gruyter.
- Kewley-Port, D., & Pisoni, D. B. (1984). Identification and discrimination of rise time: Is it categorical or noncategorical?. *The Journal of the Acoustical Society of America*, 75(4), 1168-1176.
- King, R. D. (1965). Weakly stressed vowels in Old Saxon. *Word*, 21(1), 19-39.
- Klatt, D. H. (1979). Speech perception: A model of acoustic–phonetic analysis and lexical access. *Journal of phonetics*, 7(3), 279-312.
- Klatt, D. H. (1975). Vowel lengthening is syntactically determined in a connected discourse. *Journal of phonetics*, 3(3), 129-140.
- Klatt, D. H., Cooper, W. E. (1975) Perception of segment duration in sentence contexts. In *Structure and process in speech perception*, pp. 69-89.
- Kohler, K. (1990) 'German'. *Journal of the International Phonetic Association*, 20/1:48-50.
- Kohler, K. J. (1992). Erstellen eines Textkorpus für eine phonetische Datenbank des Deutschen. In K. J. Kohler (Ed.), *Phonetisch-akustische Datenbasis des Hochdeutschen: Kieler Arbeiten zu den PHONDAT-Projekten 1989–1992*, AIPUK 26, pp. 11–39
- Kraljic, T., Brennan, S. E., Samuel, A. G. (2008). Accommodating variation: Dialects, idiolects, and speech processing. *Cognition*, 107(1), 54-81.

- Krashen, S. (1982). *Principles and Practice in Second Language Acquisition*
- Kwon, Y. K. (2011). Doing without [tense/lax] in English. *Linguistic Research*, 28(3), 605-624.
- Kyes, R. L. (1967). The evidence for i-umlaut in Old Low Franconian. *Language*, 666-673.
- Labov, W. (2001). *Principles of linguistic change Volume 2: Social factors*. LANGUAGE IN SOCIETY-OXFORD-, 29.
- Ladefoged, P., Broadbent, D. E. (1957). Information conveyed by vowels, *J. Acoust. Soc. Am.* 29, 88–104.
- Lado, R. (1957). *Linguistics across cultures*. Michigan. Ann Arbor.
- Lahiri, A., B. E. Drescher. (1999). Open syllable lengthening in West Germanic. *Language* 75: 678–719.
- Larsen-Freeman, D., Long, M.H. (1991). *An introduction to second language acquisition research*. Routledge.
- Lass, R. (1976). *English phonology and phonological theory*. Cambridge: Cambridge University Press.
- Lehiste, I. (1970). *Suprasegmentals* MIT, Cambridge, MA, pp. 18–33.
- Lehiste, I. Peterson, G.E. (1961). Transitions, Glides, and Diphthongs. *Journal of Acoustical Society of America*, 33, 268-277.
- Lennes, M. (2013). Script "collect_formant_data_from_files.praat", <http://www.helsinki.fi/~lennes/praat-scripts/>
- Lenz, A.N. (2010). *Grammar between Norm and Variation*. (Vol. 40). Peter Lang.
- Levis, J., Cortes, V. (2008). Minimal pairs in spoken corpora: Implications for pronunciation assessment and teaching. *Towards adaptive CALL: Natural language processing for diagnostic language assessment*, 197208.
- Liberman, A. M., Cooper, F. S., Shankweiler, D. P., Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychol. Rev.* 74:431– 61
- Liberman, A. M., Delattre, P., Cooper, F. S. (1952). The role of selected stimulus-variables in the perception of the unvoiced stop consonants. *Am. J. Psychol.* 65:497–516

- Liberman A. M., Delattre, P. C., Cooper, F. S., Gerstman, L. J. (1954). The role of consonant-vowel Annu. Rev. Psychol. 2004.55:149-179. Downloaded from www.annualreviews.org Access provided by University of California - Davis on 10/06/21. For personal use only. 18 Nov 2003 14:46 AR AR207-PS55-06.tex AR207-PS55-06.sgm LaTeX2e(2002/01/18) P1: GCE SPEECH PERCEPTION 177 transitions in the stop and nasal consonants. Psychol. Monogr. 68:1–13
- Liberman A. M., Delattre, P.C., Gerstman, L. J., Cooper, F. S. (1956). Tempo of frequency change as a cue for distinguishing classes of speech sounds. J. Exp. Psychol. 52:127–37
- Liberman, A. M. (1957). Some results of research on speech perception. J. Acoust. Soc. Am. 29:117–23
- Liljencrants, J., Lindblom, B. (1972). Numerical simulation of vowel quality systems: the role of perceptual contrast. Language 48: 839–862
- Lindner, G. (1976). Urteilsänderung bei Vokalkürzung. Z. Phonet. Sprachw. Kommunforsch. 29: 407-414 (1976).
- Lindsey, G. (1990). Quantity and quality in British and American vowel systems. Studies in the pronunciation of English: a commemorative volume in honour of AC Gimson, 106-118.
- Lippi-Green, R. (1997). English with an accent. Language, ideology, and discrimination in the United States. London and New York: Routledge.
<https://doi.org/10.1017/S004740450002025X>.
- Liu, C., Jin, S. H., Chen, C. T. (2014). Durations of American English Vowels by Native and Non-native Speakers: Acoustic Analyses and Perceptual Effects. Language and Speech 57.2.238 –253.
- Llompert, M., & Simonet, M. (2018). Unstressed vowel reduction across Majorcan Catalan dialects: Production and spoken word recognition. Language and speech, 61(3), 430-465.
- Llompert, M., Reinisch, E. (2018). Acoustic cues, not phonological features, drive vowel perception: Evidence from height, position and tenseness contrasts in German vowels. Journal of Phonetics 67.34-38.
- Lobanov, B.M. (1971). Classification of Russian vowels spoken by different speakers, J. Acoust. Soc. Am. 49, 606–608.
- Long, H. (2019). Learner language analysis: A case study of a Chinese EFL student. Journal of Asia TEFL, 16(3), 1013.

- Lotto, A. J., Hickok, G. S., & Holt, L. L. (2009). Reflections on mirror neurons and speech perception. *Trends in cognitive sciences*, 13(3), 110-114.
- Maack, A. (1951). Die Variation der Lautdauer deutscher Sonanten. *Zeitschrift für Phonetik und allgemeine Sprachwissenschaft* 5:287-340.
- Maack, A. (1954). Die Korrelation Akzent/Quantität. *Zeitschrift für Phonetik und allgemeine Sprachwissenschaft* 8:226-238.
- Magnuson, J. S., Nusbaum, H. C. (2007). Acoustic differences, listener expectations, and the perceptual accommodation of talker variability.
- Mangold, M. (1990). DUDEN. Aussprachewörterbuch. Mannheim etc.: Bibliographisches Institut.
- Mangold, M. (1922-2015). Duden "Aussprachewörterbuch": Wörterbuch der deutschen Standardausprache. Der Duden in zwölf Bänden.
- Martens, C., & Martens, P. (1961). *Phonetik der deutschen Sprache: praktische Aussprachelehre* (Vol. 1). M. Hueber.
- Mattingly, I.G. (2014). Modularity and the Motor Theory of Speech Perception. *Proceedings of A Conference To Honor Alvin M. Liberman*
- Maurer, D. (2016). *Acoustics of the Vowel-Preliminaries* (p. 296). Peter Lang International Academic Publishers.
- Maurer, D.; Landis, T. (1995). F0-Dependence, Number Alteration, and Non-Systematic Behaviour of the Formants in German Vowels. *International Journal of Neuroscience* 83:1-2, 25-44.
- Maye, J., Gerken, L. (2000). Learning phonemes without minimal pairs. In *Proceedings of the 24th annual Boston university conference on language development* (Vol. 2, pp. 522-533).
- McAllister, R., Flege, J. E., Piske, T. (2002). The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. *Journal of phonetics*, 30(2), 229-258.
- McClelland, J. L., Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive psychology*, 18(1), 1-86.
- McCloy, D. (2016). *phonR: Tools for Phoneticians and Phonologists*.
- McDonald, W. C. (1972). *A survey of German medieval literary patronage from Charlemagne to Maximilian I*. The Ohio State University.

- McGurk H, MacDonald J. (1976). Hearing lips and seeing voices. *Nature* 264:746–47.
- McQueen, J. M. (2005). Speech perception. In *The Handbook of Cognition* (pp. 255-275). Sage Publications.
- Menezes, V. (2013). Second language acquisition: Reconciling theories. *Open Journal of Applied Sciences*, 3(07), 404.
- Milroy, J. (2007). The ideology of the standard language. In *The Routledge Companion to Sociolinguistics*, 133-139.
- Milroy, J., Milroy, L. (1999). *Authority in language: Investigating Standard English*. London and New York: Routledge. <https://doi.org/10.4324/9780203124666>.
- Mitchell, R., Myles, F., Marsden, E. (2019). *Second language learning theories*. Routledge.
- Moosmüller, S. (2007). *Vowels in Standard Austrian German. An Acoustic-Phonetic and Phonological Analysis*. Habilitation, University of Vienna.
- Morrison, G. S. (2013). Theories of vowel inherent spectral change. *Vowel inherent spectral change*, 31-47.
- Moyer, J. (2004). *Age, accent and experience in second language acquisition*. Clevedon, Avon: Multilingual Matters.
- Murcia-Soler, M., Perez-Gimenez, F., Garcia-March, F. J. (2003). Artificial Neural Networks and Linear Discriminant Analysis: A Valuable Combination in the Selection of New Antibacterial Compounds. *J. Chem. Inf. Comput. Sci.* 44.1031-1041.
- Nearey, T. M. (1995). Speech perception as a pattern recognition. *The Journal of the Acoustical Society of America*, 97(5), 3334-3334.
- Nearey, T. M. (1989). Static, dynamic, and relational properties in vowel perception. *The Journal of the Acoustical Society of America*, 85(5), 2088-2113.
- Nearey, T.M. (1978). Phonetic feature systems for vowels. (1978): 4792-4792.
- Nearey, T. M., Assmann, P. F. (1986). Modeling the role of inherent spectral change in vowel identification. *Journal of the Acoustical Society of America*, 80, 1297-308.
- Newton, R. P. (2019). The Nature of Vowel Length. In *Vowel undersong* (pp. 141-156). De Gruyter Mouton.

- Nishi, K., Strange, W., Akahane-Yamada, R., Kubo, R., Trent-Brown, S. A. (2008). Acoustic and perceptual similarity of Japanese and American English vowels. *The journal of the Acoustical Society of America*, 124(1), 576-588.
- Nittrouer, S. (2000). Learning to apprehend phonetic structure from the speech signal: the hows and whys. Paper presented at the VIIIth Meeting of the International Clinical Phonetics and Linguistics Association, Queen Margaret University College, Edinburgh.
- Nowak, M. A., Komarova, N. L., Niyogi, P. (2001). Evolution of universal grammar. *Science*, 291(5501), 114-118.
- Nübling, D. (2006). *Historische Sprachwissenschaft des Deutschen. Eine Einführung in die Prinzipien des Sprachwandels.*
- Nusbaum, H. C., Magnuson, J. S (1997). Talker normalization: Phonetic constancy as a cognitive process. *Talker variability in speech processing*, 109-132.
- Nusbaum, H. C., Morin, T. M. (1992). Paying attention to differences among talkers. In Y. Tohkura, Y. Sagisaka, & E. Vatikiotis-Bateson (Eds.), *Speech perception, production, and linguistic structure.* (pp. 113-134). Tokyo: OHM Publishing Company.
- O'Brien, M. G., & Fagan, S. M. (2016). *German phonetics and phonology: Theory and practice.* Yale University Press.
- O'brien, M. G., & Smith, L. C. (2010). Role of first language dialect in the production of second language German vowels.
- Pallotti, G. (2017). Applying the interlanguage approach to language teaching. *International review of applied linguistics in language teaching*, 55(4), 393-412.
- Parker, E. M., Diehl, R. L. (1984). "Identifying vowels in CVC syllables: Effects of inserting silence and noise," *Percept. Psychophys.* 36, 369–380
- Pätzold, M., Simpson, A. (1997) Acoustic analysis of German vowels in the Kiel Corpus of Read Speech. *Institut für Phonetik und digitale Sprachverarbeitung Universität Kiel, Arbeitsberichte (AIPUK) 32.*
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, 12, 2825-2830.
- Penzl, H. (1971). *Lautsystem und Lautwandel in den althochdeutschen Dialekten.* Max Hueber Verlag, München.

- Penzl, H. (1949). Umlaut and secondary umlaut in Old High German. *Language*, 25(3), 223-240.
- Perkins, M. (1983). *Sensing the world*. Indianapolis, IN: Hackett.
- Péry-Woodley, M. P. (1990). Contrasting discourses: contrastive analysis and a discourse approach to writing. *Language Teaching*, 23(3), 143-151.
- Peterson, G. E. (1961). Parameters of vowel quality. *Journal of Speech and Hearing Research*, 4(1), 10-29.
- Peterson, G.E., Barney, H.L. (1952). Control methods used in a study of the vowels. *J. Acoust. Soc. Am.* 24, 175–184
- Peterson, G.E. , Lehiste, I. (1960). Duration of syllable nuclei in English. *Journal of the Acoustic Society of America*, 32, 693-702.
- Pierrehumbert, J., Hay, J., Beckman, M. (1998). Speech perception, wellformedness & lexical frequency. In the 6th Conference on Laboratory Phonology, York, England.
- Polka, L. (1995). Linguistic influences in adult perception of non native vowel contrasts. *The Journal of the Acoustical Society of America*, 97(2), 1286-1296.
- Predeck, K., Block, A., Arnett, C. (2021). Problematic phonemes” and German/ε:: An acoustic analysis. *Proceedings of the Linguistic Society of America*, 6(1), 280-287.
- Ramers, K.H. (1988). *Vokalquantität und -qualität im Deutschen*. Tübingen: Niemeyer (= linguistische Arbeiten 213).
- Rauch, I. (2017). *The old high German diphthongization*. De Gruyter Mouton.
- Renwick, M. E. (2014). *The Phonetics and Phonology of Contrast*. In *The Phonetics and Phonology of Contrast*. De Gruyter Mouton.
- Riad, T. (1995). The quantity shift in Germanic: A typology. *Amsterdamer beiträge zur älteren germanistik*, 42, 159.
- Riad, T. (1992). *Structures in Germanic prosody*. Doctoral dissertation, University of Stockholm.
- Rogalski, C., Love, T., Driscoll, D., Anderson, S.W (2011). Are mirror neurons the basis of speech perception? Evidence from five cases with damage to the purported human mirror system
- Rosner, B. S., & Pickering, J. B. (1994). *Vowel perception and production* (Vol. 23).

OUP Oxford.

- Russ, C. (2002). *The German language today: A linguistic introduction*. Routledge.
- Rustipa, K. (2011). Contrastive analysis, error analysis, interlanguage and the implication to language teaching. *Ragam Jurnal Pengembangan Humaniora*, 11(1), 16-22.
- Salmons, J. (2018). *A history of German: What the past reveals about today's language*. Oxford University Press.
- Samuel, A. G. (1977). The effect of discrimination training on speech perception: Noncategorical perception. *Perception & Psychophysics*, 22(4), 321-330.
- Schmid, C., & Moosmüller, S. (2017). An acoustic comparison between stressed and unstressed vowels in Standard Austrian German and Standard German German. Sylvia Moosmüller, Carolin Schmid & Manfred Sellner (Hgg.), *Phonetik in und über Österreich*, 45-59.
- Schoormann, H. E., Heeringa, W., Peters, J. (2019). Standard German vowel productions by monolingual and trilingual speakers. *International Journal of Bilingualism*, 23(1), 138–156. <https://doi.org/10.1177/1367006917711593>
- Schultheiss, L. K. (2008). *Cross-language perception of German vowels by speakers of American English*. Brigham Young University.
- Schwartz, G. (2021). The phonology of vowel VISC-osity—acoustic evidence and representational implications. *Glossa: a journal of general linguistics*, 6(1).
- Scobbie, J. (1998). 'Interactions between the Acquisition of Phonetics and Phonology'. In Gruber, K., Higgins, D., Olsen, K. and Wysochi, T. (eds.). *Papers from the 34th Annual Meeting of the Chicago Linguistic Society, II*. Chicago: Chicago Linguistic Society.
- Selinker, L. (1972). "Interlanguage." (1972): 209-232.
- Sendlmeier, W., Seebode, J. (2006) Formantkarten des deutschen Vokalsystems. <www.kw.tu-berlin.de/fileadmin/a01311100/Formantkarten_des_deutschen_Vokalsystems_01.pdf> Last accessed 14.05.2012.
- Sendlmeier, W.F. (1981). Der Einfluss von Qualität und Quantität auf die Perzeption betonter Vokale des Deutschen, *Phonetica* 38, 291–308.
- Shi, L., Feldman, N. H., Griffiths, T. L. (2008). Performing Bayesian inference with exemplar models. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 30, No. 30).

- Shi, L., Feldman, N. H., Griffiths, T. L. (2008). Performing Bayesian Inference with Exemplar Models. In Proceedings of the Annual Meeting of the Cognitive Science Society (Vol. 30, No. 30).
- Shi, L., Feldman, N. H., Griffiths, T. L., Sanborn, A. (2010). Exemplar models as a mechanism for performing Bayesian inference. *Psychonomic bulletin & review*, 17(4), 443-464.
- Siebs, T. (1969). Deutsche Aussprache. Reine und gemäßigte Hochlautung mit Aussprachewörterbuch, 19.
- Siebs, T. (2020). Die Laute der deutschen Bühnensprache. In *Deutsche Bühnenaussprache. Hochsprache* (pp. 23-84). De Gruyter.
- Siebs, T. (1949). Umlaut and secondary umlaut in Old High German. *Language*, 25(3), 223-240.
- Skinner, B. F. (1957). *Verbal behavior*. New York: Appleton-Century-Crofts, 1957.
- Slabakova, R. (2017). The scalpel model of third language acquisition. *International Journal of Bilingualism*, 21(6), 651-665.
- Slawson, A. W. (1968). Vowel quality and musical timbre as functions of spectrum envelope and fundamental frequency. *The Journal of the Acoustical Society of America*, 43(1), 87-101.
- Spolsky, B. (1989). Communicative competence, language proficiency, and beyond. *Applied Linguistics*, 10(2), 138-156.
- Steinlen, A. K. (2005). *The Influence of Consonants on Native and Non-native Vowel Production: A Cross-linguistic Study*
- Stevenson, P. (2002). *Language and German disunity: a sociolinguistic history of East and West in Germany, 1945-2000*. Oxford University Press on Demand.
- Strange, W. (2007). Cross-language similarity of vowels. Theoretical and methodological issues. *Language Experience in Second Language Speech Learning: In honor of James Emil Flege* 35-55.
- Strange, W. (1987). Information for vowels in formant transitions. *Journal of Memory and Language* 26.5.550-557.
- Strange, W. (1989). 'Dynamic specification of coarticulated vowels spoken in sentence context,' *J. Acoust. Soc. Am.* 85, 2135–2153.

- Strange, W., Bohn, O. S., Nishi, K., Trent, S. A. (2005). Contextual variation in the acoustic and perceptual similarity of North German and American English vowels. *The Journal of the Acoustical Society of America*, 118(3), 1751-1762.
- Strange, W., Weber, A., Levy, E. S., Shafiro, V., Hisagi, M., Nishi, K. (2007). Acoustic variability within and across German, French, and American English vowels: Phonetic context effects. *The Journal of the Acoustical Society of America*, 122(2), 1111-1129.
- Strange, W.; Bohn, O.S. (1998). Dynamic specification of coarticulated German vowels: Perceptual and acoustical studies. *The Journal of the Acoustical Society of America* 104.488.
- Strange, W.; Bohn, O.S.; Trent, S.; Nishi, K. (2004). Acoustic and perceptual similarity of North German and American English vowels. *The Journal of the Acoustical Society of America* 115.1791.
- Szulc, A. (1987). *Historische Phonologie des Deutschen*. Max Niemeyer Verlag.
- Tkaczyk, V. (2017). Hochsprache im Ohr. In *Wissensgeschichte des Hörens in der Moderne* (pp. 123-152). De Gruyter.
- Tomaschek, F. (2013). Behavioral and neural correlates of vowel length in German and of its interaction with the tense/lax contrast.
- Tomaschek, F., Truckenbrodt, H., Hertrich, I. (2015). Discrimination sensitivities and identification patterns of vowel quality and duration in German /u/ and /o/ instances. Lang.
- Tomaschek, F., Truckenbrodt, H., Hertrich, I. (2013). Neural processing of acoustic duration and phonological German vowel length: Time courses of evoked fields in response to speech and nonspeech signals.
- Tomaschek, F., Truckenbrodt, H., Hertrich, I. (2011). Processing German Vowel Quantity: Categorical Perception or Perceptual Magnet Effect?
- Twaddell, W. F. (1938). A note on Old High German umlaut. *Monatshefte für deutschen Unterricht*, 177-181.
- Ungeheuer, G. (1969). Das Phonemsystem der deutschen Hochlautung, in Siebs (1969), 27-42.
- van Santen, J. (1992). Contextual effects on vowel duration. *Speech Communication*, 11, 513-46.
- Vennemann, T. (2000). From quantity to syllable cuts: On so-called lengthening in the

- Germanic languages. *Italian Journal of Linguistics / Rivista di Linguistica* 12: 251–282.
- Verbrugge, R., Rakerd, B. (1986). "Evidence of talker-independent information for vowels," *Lang. Speech* 29, 39–57.
- Vernon, J.P. (1976). La neutralisation de l'opposition de durée et de timbre des voyelles de l'allemand. *Cahier de l'Allemand* 11:110-126.
- von Essen, O. (1979). *Allgemeine und angewandte Phonetik*, 5. Auflage. Akademie Verlag Berlin.
- Voyles, J., Rauch, I., Carr, G. F., & Kyes, R. L. (1992). On Old High German i-umlaut. *On Germanic Linguistics: Issues and Methods*. Berlin: Mouton de Gruyter, 365-377.
- Voyles, J.B. (1976). Old High German Umlaut. *Zeitschrift für vergleichende Sprachforschung*, 90(1./2. H), 271-289.
- Wardhough, R. (1974). *Topics in Applied Linguistics*.
- Waterman, J. T. (1966). *A history of the German language: with special reference to the cultural and social forces that shaped the standard literary language*. Waveland PressInc.
- Watson, C. S., Kewley-Port, D., Foyle, D. C. (1985). Temporal acuity for speech and nonspeech sounds: The role of stimulus uncertainty. *The Journal of the Acoustical Society of America*, 77(S1), S27-S27.
- Weiss, R. (1976). The perception of vowel length and quality in German: An experimental-phonetic investigation (Vol. 20). Buske.
- Weiss, R. (1977). The phonemic significance of the phonetic factors of vowel length and quality in German. In Dressler, W.U. and Pfeiffer, O.E. (eds.) *Phonologica 1976*. Innsbruck: Innsbrucker Beiträge zur Sprachwissenschaft. 271-276.
- Werner, O. (1972). *Phonemik des Deutschen*. JB Metzler.
- White, L. (2003). *Second language acquisition and universal grammar*. Cambridge University Press.
- Wiese, R. (1996). *The Phonology of German*. Oxford University Press on Demand.
- Wiese, R. (1996). Phonological versus morphological rules: On German umlaut and ablaut. *Journal of Linguistics*, 32(1), 113-135.
- Wiese, R. (1996). *The Phonology of German* Clarendon, Oxford

Wrede, B., Fink, G., Sagerer, G. (2000). Influence of duration on static and dynamic properties of German vowels in spontaneous speech. 6th International Conference on Spoken Language Processing.

Wright, R. (2004). A review of perceptual cues and cue robustness. In B. S. Bronson (Ed.), *Phonetically based phonology* (pp. 34– 57). Cambridge: Cambridge University Press.

Wright, R., Frisch, S., Pisoni, D.B. (1999). "Speech perception." *Wiley encyclopedia of electrical and electronics engineering* 20 (1999): 175-195.

Zsiga, E.C. (2013). *The Sounds of Language. An Introduction to Phonetics and Phonology*.

Appendix

1. Spectral continua

/a:/

Step	F1	F2
1	575	1286
2	618	1250
3	662	1215
4	706	1180
5	749	1145

/e:/

Step	F1	F2
1	432	1750
2	417	1822
3	401	1894
4	384	1966
5	369	2039

/i:/

Step	F1	F2
1	314	1884
2	355	1987

3	396	2090
4	437	2193
5	480	2298

/o:/

Step	F1	F2
1	512	1084
2	483	1076
3	454	1068
4	425	1060
5	396	1052

/ø:/

Step	F1	F2
1	521	1567
2	481	1552
3	442	1538
4	403	1526
5	364	1513

/u:/

Step	F1	F2
------	----	----

1	343	1197
2	347	1271
3	351	1345
4	355	1419
5	361	1492

/y:/

Step	F1	F2
1	372	1576
2	348	1613
3	325	1650
4	302	1687
5	279	1725

/ɛ:/

Step	F1	F2
1	483	1673
2	482.5	1689
3	482	1705
4	481.5	1721
5	481	1738

2. Duration Continua

/ɛ:/

Step	Duration (in ms)
1	437
2	413
3	389
4	365
5	341

/a:/

Step	Duration (in ms)
1	413
2	381
3	350
4	319
5	288

/y:/

Step	Duration (in ms)
1	341
2	324

3	307
4	290
5	273

/u:/

Step	Duration (in ms)
1	291
2	284
3	279
4	274
5	269

/o:/

Step	Duration (in ms)
1	409
2	369
3	331
4	293
5	255

/i:/

Step	Duration (in ms)
1	314
2	294
3	275
4	256
5	237

/e:/

Step	Duration (in ms)
1	440
2	399
3	357
4	315
5	273

/ø:/

Step	Duration (in ms)
1	405
2	364
3	325
4	286

5	247
---	-----

3. Model outputs for linear mixed-effects models run on F1 and F2, and duration for long/short vowel pairs (lmer syntax: `acousticMeasure ~ vowel + (1|participant)`)

/ɛ:/-/ɛ/

	<i>Est</i>	<i>Std. Err.</i>	<i>t</i>	<i>p</i>
(Intercept)	-0.83	0.06	-12.27	<0.001
F1	0.005	0.01	0.35	0.72

	<i>Est</i>	<i>Std. Err.</i>	<i>t</i>	<i>p</i>
(Intercept)	0.37	0.02	13.9	<0.001
F2	0.03	0.005	6.33	<0.001

	<i>Est</i>	<i>Std. Err.</i>	<i>t</i>	<i>p</i>
(Intercept)	0.32	0.01	24.56	<0.001
duration	0.07	0.005	13.54	<0.001

/a:/-/a/

	<i>Est</i>	<i>Std. Err.</i>	<i>t</i>	<i>p</i>
(Intercept)	-0.60	0.06	-9.66	<0.001
F1	0.17	0.01	13.58	<0.001

	<i>Est</i>	<i>Std. Err.</i>	<i>t</i>	<i>p</i>
(Intercept)	0.05	0.03	1.61	0.14
F2	-0.05	0.007	-8.37	<0.001

	<i>Est</i>	<i>Std. Err.</i>	<i>t</i>	<i>p</i>
(Intercept)	0.29	0.01	23.99	<0.001
duration	0.06	0.03	17.48	<0.001

/e:/-/e/

	<i>Est</i>	<i>Std. Err.</i>	<i>t</i>	<i>p</i>
(Intercept)	-1.004	0.04	-21.18	<0.001
F1	-0.34	0.01	-30.07	<0.001

	<i>Est</i>	<i>Std. Err.</i>	<i>t</i>	<i>p</i>
(Intercept)	0.38	0.04	9.39	<0.001
F2	0.17	0.007	23.64	<0.001

	<i>Est</i>	<i>Std. Err.</i>	<i>t</i>	<i>p</i>
(Intercept)	0.28	0.01	17.18	<0.001
duration	0.09	0.008	11.27	<0.001

/i:/-/i/

	<i>Est</i>	<i>Std. Err.</i>	<i>t</i>	<i>p</i>
(Intercept)	-1.56	0.04	-34.34	<0.001
F1	0.20	0.01	14.61	<0.001

	<i>Est</i>	<i>Std. Err.</i>	<i>t</i>	<i>p</i>
(Intercept)	0.61	0.03	16.34	<0.001
F2	-0.19	0.004	-43.70	<0.001

	<i>Est</i>	<i>Std. Err.</i>	<i>t</i>	<i>p</i>
(Intercept)	0.35	0.01	27.76	<0.001
duration	-0.10	0.006	-15.69	<0.001

/o:/-/o/

	<i>Est</i>	<i>Std. Err.</i>	<i>t</i>	<i>p</i>
(Intercept)	-1.18	0.04	-26.58	<0.001
F1	0.35	0.007	44.13	<0.001

	<i>Est</i>	<i>Std. Err.</i>	<i>t</i>	<i>p</i>
(Intercept)	-0.54	0.02	-19.16	<0.001
F2	0.53	0.01	27.31	<0.001

	<i>Est</i>	<i>Std. Err.</i>	<i>t</i>	<i>p</i>
(Intercept)	0.38	0.01	26.37	<0.001
duration	-0.10	0.003	-31.21	<0.001

/u:/-/ʊ/

	<i>Est</i>	<i>Std. Err.</i>	<i>t</i>	<i>p</i>
(Intercept)	-1.11	0.09	-11.48	<0.001
F1	-0.04	0.01	-3.41	<0.001

	<i>Est</i>	<i>Std. Err.</i>	<i>t</i>	<i>p</i>
(Intercept)	-0.01	0.13	-0.08	0.93
F2	0.004	0.03	0.13	0.88

	<i>Est</i>	<i>Std. Err.</i>	<i>t</i>	<i>p</i>
(Intercept)	0.32	0.01	21.67	<0.001
duration	-0.02	0.003	-9.33	<0.001

/y:/-/ʏ/

	<i>Est</i>	<i>Std. Err.</i>	<i>t</i>	<i>p</i>
(Intercept)	-1.46	0.06	-23.15	<0.001
F1	0.19	0.01	15.32	<0.001

	<i>Est</i>	<i>Std. Err.</i>	<i>t</i>	<i>p</i>
(Intercept)	0.33	0.02	11.43	<0.001
F2	-0.06	0.009	-7.47	<0.001

	<i>Est</i>	<i>Std. Err.</i>	<i>t</i>	<i>p</i>
(Intercept)	0.34	0.02	14.77	<0.001

duration	-0.06	0.004	-15.36	<0.001
-----------------	--------------	--------------	---------------	------------------

/ø:/-/œ/

	<i>Est</i>	<i>Std. Err.</i>	<i>t</i>	<i>p</i>
(Intercept)	-1.26	0.04	-27.57	<0.001
F1	0.38	0.01	26.88	<0.001

	<i>Est</i>	<i>Std. Err.</i>	<i>t</i>	<i>p</i>
(Intercept)	0.19	0.03	5.51	<0.001
F2	0.03	0.01	2.55	0.01

	<i>Est</i>	<i>Std. Err.</i>	<i>t</i>	<i>p</i>
(Intercept)	0.37	0.01	28.78	<0.001
duration	-0.09	0.004	-20.20	<0.001

4. Demographic Tables

Experiment One German: Background information for German speakers

ID	Gender	Place of Residence	Age	Other languages
1	m	NRW	56	-
2	m	Swabia	50	English
3	w	NRW	44	-
4	m	NRW	59	English, French
5	w	NRW	40	-
6	m	NRW	44	-
7	m	Baden-Württemberg	32	-

8	m	NRW	71	-
9	m	NRW	49	-
10	w	Hesse	56	English
11	w	Bremen	55	-
12	m	NRW	56	-
13	w	NRW	58	-
14	w	NRW	64	-
15	w	NRW	30	-
16	w	NRW	55	-
17	w	NRW	55	-
18	w	Saxony-Anhalt	40	-
19	m	NRW	50	-
20	m	NRW	55	-
21	w	NRW	30	-

Experiment One English: Background information for English speakers

ID	Gender	Place of Residence	Age	Other languages
84376	f	CA	21	Chinese
86043	f	CA	20	Cantonese
87839	f	CA	20	Cantonese
92084	non-binary	CA	19	Telugu
98090	m	CA	19	-
98109	m	CA	19	Spanish
99974	m	CA	20	-
99984	f	CA	19	Arabic
100044	m	CA	19	Tonga
100142	f	CA	43	Gullah
100193	f	CA	27	Spanish, French

Experiment two: Background information for German speakers

ID	Gender	Age	Other languages
R_1prR00HgyHxCrKi	m	57	-
R_1cSvKaboYXpkcC9	m	30	English
R_2e4pG6zvC8OmVPd	f	30	English
R_3D5g2sIkWduCB3D	f	24	-
R_6PuYcnIOxj1iaB	m	54	English
R_1Kdy7YrI9gfFO9r	m	23	English, Italian
R_2X0adVvhQXPmc9w	f	67	English
R_9BK2LRIO8Iq8nC1	f	26	English
R_3njyRvAGWggfAnI	f	30	-
R_1kFd7YlwpyoZwyS	f	30	-
R_zZ7UmjdnHQTbwBz	f	19	English
R_9EW0rgyAzNnneIV	m	39	English
R_uxntH3xYUVQuqZj	f	23	English
R_3oGuD5Mkmnnq1yq	f	21	English
R_2ffJWyZYGeUKLby	f	24	English
R_2zw0Q00Nh5RFS1P	f	18	English, Spanish
R_3qxPpaprVDu3r9e	m	25	English
R_PUPZfyMisr3VGYV	f	19	English
R_1lo7jucj4TBggY7	f	41	English
R_3HifyJKipRPEdva	f	19	English, a little French
R_2UhhjakVhBfOp6O5	f	24	-
R_1i3bCktZekbZiZQ	f	20	-
R_1dMgYKDIO70y1FA	f	27	English
R_3g7V3SL8YH5ZdeY	f	23	English, Spanish, Italian
R_sHeN7FJQ84th9ex	m	18	English, Russian
R_2atPZo8LVSxrSam	f	20	English
R_4Ib58cvHwHhYjQZ	f	34	English
R_vAnAwKQRXKrHY2J	f	20	-
R_1IZ8K0Hgy0goqmT	f	29	English
R_2R98ocFx7FLiby5	f	22	-

R_2BA18SkyDQpcTEi	f	24	English
R_3NCnIEMYeZYwDAq	f	21	Turkish, English
R_10Re0jsouYeaDK	non-binary	23	English
R_3nuyxEsnh1DYauf	m	20	English
R_1FzE51ktgY9tRZL	f	19	English
R_3fiVMFCU2P6RCJH	m	21	English
R_0lyGN0Ry47S6d3P	non-binary	28	-
R_31ocvQKS7NzmxPE	f	45	Turkish
R_e5akOlaOlKOUlyd	f	18	English
R_1rPr1eWxilswMnx	f	21	French
R_3DwPDok185QnwSq	f	18	English
R_2e2TALs4XUflImi	f	24	English
R_1hHnANti7DRoXWp	m	24	Turkish, English
R_2akra7XfUcd98Vs	m	28	English
R_1Qc431yo7zboK4u	f	35	-
R_3dZbZqOxgcyw9yq	f	23	English, Russian
R_1pLNBowTbuNxPZB	non-binary	19	English
R_2uBNjRXNp09miJm	f	25	English, French
R_3ffxZpLw1ECaQ92	m	36	English
R_2qeA1HZeyssltn3	f	28	Spanish
R_3qW1dowZpMayVxu	m	40	Italian, English, Spanish
R_2rGvq033xOzLwjn	f	23	English
R_O89wqBgY3w7oYNP	m	33	English
R_3J2AvTTCpmwyGau	m	24	English
R_3fZY84botY7hviP	f	47	English, French
R_Z3rde0qN4W4Vy93	f	29	English, Russian
R_1eEPEamkKdlawLw	f	19	English, French
R_3fYJSCwXXuJr7HV	f	30	English, French
R_exhDDDSCy5MibJf	f	35	English, Hungarian
R_1jlbZV4l1aAgJvZ	f	18	English
R_8dE5zakOJhrwTbr	f	25	English
R_240OFYacLD5Skfw	m	35	English, Japanese
R_2dhVnibDILcJyJC	m	45	English
R_2zZsk8OZHn0w2Ee	m	56	English, French, Welsh

R_1MS9WVqoCLWL6Tq	f	42	English
-------------------	---	----	---------

Experiment three: Background information for German speakers

ID	Gender	Age	Other languages
R_2BhUS3pyf0IEW3M	m	57	-
R_Zb0BanYZZRqIUhj	m	54	English
R_w5FUP640IJgLvz	m	23	English, Italian
R_3ER3rwwDt45hEG0	f	30	-
R_4GhIlbcb5S3UMtX	f	19	English
R_etjxjTWE2Rt4WsN	f	23	English
R_31FLIAADrmBCJCP	f	21	English
R_1r6MvT8EezmjYaYY	f	41	English
R_2Ei5HcLD3xxZ7EQ	f	19	English, French
R_C7hlyB1LQHHY2Qh	f	24	-
R_3vYckyCCFqiFhm1	f	20	-
R_8wBoKWWllgHXMvn	f	27	English
R_2PhCpuEcY7Wk2SJ	f	24	English
R_Tn2ZU38dKrpaiEN	f	20	-
R_24r6Zu2pgjacROx	non-binary	28	-
R_2xVkp6lyPaw6Ebh	f	35	-
R_3lSnLdnd8tvC6L4	f	21	English, French
R_1DTCVF73jZiA4rT	m	28	English
R_294VDTuGDC1GUOE	f	23	English, Russian
R_1E6zaINUfXb8wdb	f	28	Spanish
R_2uxRtjQiUUd2QHb	f	23	English
R_1lBfscgaHcX8768	m	36	English
R_W226ja9Jvn46ovv	m	24	English
R_1gC31OjV8l16qSR	f	47	English, French
R_2YCU5trZ7xlEqiq	f	29	English, Russian
R_cBxtepzUDeSLRv	f	19	English
R_10Q2NDKyMIUwzjo	f	25	English
R_3PYxlp5Vq7ax2Q8	m	26	English, Greek, French, Italian

R_2scdfaeyjduxgdU	f	45	-
R_0c5I1U0liOjBhyF	f	32	English
R_1q8QcBDZH34Ibtd	m	72	English, Dutch
R_2trQot1Yqwps5TF	f	33	English, Finnish
R_1kYoXqhSCcsO2Z3	f	54	English
R_2BmMFXrlmpx3jmX	f	30	English
R_3NMXj65XPn0rG9j	f	56	-
R_29fTYBRBnvgr3P	f	31	English
R_2c89oDr4JWEK1QZ	f	39	-
R_AKHp5uef8OZ15NT	m	51	-
R_3JapVeyrvkoAAfr	non-binary	37	English, Italian, Swedish
R_33yb3mxXXdJu9oO	m	60	English
R_1EYyBXh6B5kNiyK	f	65	English
R_10Jc6aMDfEsmrA9	m	31	English
R_2rpBGKaxkFIIGV3	f	30	English
R_1Ov0Dhdot0q41MG	m	53	English
R_VI67VFYIV64VYVB	f	27	English
R_3gYkDha7nkoHg6j	m	57	English
R_21tAm2zGF3Ga08I	m	57	-
R_2b2YUT4g6S4Tng1	f	30	English
R_1AAye4sdvCuQ6vn	m	58	-
R_1ikTY1HHr1d1932	m	17	English
R_RxhwecwOQfMp9Nn	f	53	-
R_2zGgKzNmES1GJ5U	f	50	-
R_2zGavnMcvCuui4Q	f	32	English, Swedish, Norwegian, Italian
R_2zYcCarPbJvv5XR	m	51	-
R_2wb4k5vhcxYaMvN	f	21	English
R_301aEKP5UCINUqv	m	48	Czech
R_XUmRfa1jWEROUcd	m	55	English

Experiment four: Background information for German speakers

ID	Gender	Age	Other languages
R_3xehOEyhtRQ8zrX	m	57	-
R_2Ylaup6j7RqKqSB	m	54	English
R_30dAnKkkyBMjcDA	f	30	-
R_WCiDqw0hZHqRFD3	f	19	English
R_1rAEboyHpz07Bty	f	23	English
R_YYp3LXQSWJXcfBf	f	41	English
R_UL4RzvQg6QKrxVn	f	24	-
R_2WAwqDhPFvRJ7tQ	f	20	-
R_3O6SE2UF8hz5xiN	f	33	English
R_3gXwMySg6ewFN6Z	m	18	English, Russian
R_XTDJUIcSdDBsaD7	non-binary	23	-
R_31LBqVHprrhZWDE	non-binary	28	-
R_2uDC7yYFQve3z69	f	21	French, English
R_QbqVO9Av641dssh	m	28	English
R_1DP60daUwKqKoEa	m	44	Dutch
R_OHSmwJhciA9nEsN	f	28	Spanish
R_3k1DQL54bU1ACPG	f	23	English
R_vpKBrvhAa9eYRhv	m	24	English
R_2r1PI2R8nr2kO	f	47	English, French
R_2415eMJwpF1CCf5	f	25	English
R_2qaPF2y5Qzk9mZ9	m	36	English
R_3NQjrvnGCcb7JBG	f	19	English
R_2c636bqMgtkt7QW	f	32	English
R_1ojsuYbxLDHMaQR	f	54	English
R_yVJoj231KCo9jMd	m	72	Dutch, English
R_20MAMpjDUtub0dW	f	33	English, Finnish
R_1mdQszW5f7JXaQa	f	30	English
R_4SFIPMbiTWWk4sp	f	33	-
R_O71CHmCeNpr5uud	f	31	English
R_1q9CKljKmAdYnlm	f	39	-
R_0JTObgSurT6x9p7	m	51	-
R_0diGcuQTOtQvmLf	f	56	-

R_XyWvDU05Hq0BAcx	m	60	English
R_vitNzEqSUc7kzU5	f	65	English
R_1MQzmZI04I3lrqi	m	30	English
R_274mA7kl4TfUc2G	f	19	English
R_SCoS6Uf6t8ZajTj	m	31	English
R_1OKduo39UFeuxzm	m	28	-
R_3PcTyAtdz5iW3Kr	f	30	English
R_2VQEdaFHkyGKlgX	non-binary	21	English, Russian
R_qPGSc4oBcHcvFbX	f	19	English
R_SHRbhoJ4JUvKWEp	f	21	English
R_cHAjzAf02bRhHk5	f	21	English
R_2wuj3C5wYiyUg2S	f	22	English, Dutch
R_8ob1erAtKVxiDTz	f	22	English
R_27lh0lWWjuFhyHw	m	22	English
R_0l1eodoFbsGtWp3	f	19	English
R_22EaSOi3R18B2YT	non-binary	19	English, French
R_1FDStPlxgKGwzPm	f	42	English, Spanish, Hungarian
R_1oolDawUMutAoeI	m	23	English
R_1mtmoimHaeAbpwF	m	19	English, Dutch
R_30o4rfLerL6ZHU2	m	22	English
R_6fG92fjCFfViDTz	m	20	English
R_3R9DyrPIAgNOUu2	f	20	English
R_eLOcv92Ny3VsG65	f	16	English
R_RJ1hSUjVaBRvle1	f	19	English
R_XvWa9QPyOsasyIj	f	20	English
R_3rHFtInXuNcZdHn	m	20	English
R_2D8km4DKbRf5BmN	m	22	English
R_00X3HZgnRG7OM9z	m	53	English