ABSTRACT

Title of dissertation: BAYESIAN MODEL OF CATEGORICAL EFFECTS IN L1 AND L2 SPEECH PERCEPTION

Yakov Kronrod, Doctor of Philosophy, 2014

Dissertation directed by: Professor Naomi Feldman
Department of Linguistics

In this dissertation I present a model that captures categorical effects in both first language (L1) and second language (L2) speech perception. In L1 perception, categorical effects range between extremely strong for consonants to nearly continuous perception of vowels. I treat the problem of speech perception as a statistical inference problem and by quantifying categoricity I obtain a unified model of both strong and weak categorical effects. In this optimal inference mechanism, the listener uses their knowledge of categories and the acoustics of the signal to infer the intended productions of the speaker. The model splits up speech variability into meaningful category variance and perceptual noise variance. The ratio of these two variances, which I call τ, directly correlates with the degree of categorical effects for a given phoneme or continuum. By fitting the model to behavioral data from different phonemes, I show how a single parametric quantitative variation can lead to the different degrees of categorical effects seen in perception experiments with different phonemes. In L2 perception, L1 categories have been shown to exert an effect on how L2 sounds are identified and how well the listener is able to discriminate them.

Various models have been developed to relate the state of L1 categories with both the initial and eventual ability to process the L2. These models largely lacked a formalized metric to measure perceptual distance, a means of making a-priori predictions of behavior for a new contrast, and a way of describing non-discrete gradient effects. In the second part of my dissertation, I apply the same computational model that I used to unify L1 categorical effects to examining L2 perception. I show that we can use the model to make the same type of predictions as other SLA models, but also provide a quantitative framework while formalizing all measures of similarity and bias. Further, I show how using this model to consider L2 learners at different stages of development we can track specific parameters of categories as they change over time, giving us a look into the actual process of L2 category development.

Keywords: perceptual magnet effect, categorical perception, speech perception, Bayesian inference, rational analysis

BAYESIAN MODEL OF CATEGORICAL EFFECTS IN L1 AND L2 SPEECH PERCEPTION

by

Yakov Kronrod

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2014

Advisory Committee:
Professor Naomi Feldman, Chair/Advisor
Professor William Idsardi, Co-Advisor
Professor Jeffrey Lidz
Professor Kira Gor
Professor Rochelle Newman

Dedication

To my grandfathers, Alexander Pavlovich Lavut, and Alexander Semyonavich Kronrod. Gone but not forgotten. My heroes. My idols. My inspiration. Family rocks. Trendsetters. Seekers of truth and justice.

Acknowledgments

Like David Copperfield, the only way I can begin this acknowledgement is at the beginning. I was born to Tatyana Lavut and Vladimir Kronrod: my parents, my close friends, my best supporters, my greatest cheerleaders, and the two people who are singlehandedly responsible for most of who I am. There is no way to thank them enough. They raised me to respect life, truth, and justice. They brought me to this country so that I can pursue a better life. They accepted me no matter who I was or what I did in life, and they supported me every step of the way. I do not know anybody blessed with two such parents, which is a testament that all my friends who have met them over the years can agree with.

Great as they are, they could teach me no linguistics. That debt of gratitude I owe elsewhere. There are three people who stand out above and beyond all others at my time at UMD who helped shape me as a linguist. First, Naomi Feldman, my dissertation advisor. Naomi helped me navigate the road from machine translation to computational phonology. She helped me take a class project based on a small study with fricatives and a computational model she developed in graduate school and nurture it over three years into multiple conference presentations, a journal article, and ultimately my dissertation. It was not easy to have an advisor younger than you and to be their first student, but together we grew over the years and eventually found a common language and common approach to some very interesting theoretical questions. I am a better researcher and person for having stumbled along the way, thanks for helping me regain my balance. Then there is Philip Resnik,

my first advisor, and as far as I am concerned, the only reason I ever ended up at Maryland. I have never met a more enthusiastic and supportive advisor or professor. While our work together ended early, his support never wavered and in the end he helped me find my place in the department and in academia. Finally, there is Howard Lasnik. When I came to Maryland, Howard made me feel welcome and taught me most of what I know about syntax. I'll never forget the weekly games of ping pong, the discussions in the office, the chats as he came around to visit with the students, and the endless wisdom of what it means to be a devoted teacher.

Of course there were many others who helped me along the way, directly and indirectly making this dissertation possible. Robert Futrelle, working with you on BioNLP at Northeastern set in motion the series of events that would bring me to computational linguistics, I can't thank you enough. Bill Idsardi was there for my first research project I conducted in my first year at UMD and he was there when I put the finishing touches on my dissertation. In between, through highs and lows, he provided encouragement, guidance, and, sometimes, financial support, to help make the research possible. His wealth of knowledge in statistics, experimental methods, cognitive science, and basically anything else you ask him about never seizes to amaze me. Colin, you taught me what it means to built a diverse and cohesive interdisciplinary program, how to build consensus, and how to break down walls and build bridges. Thank you for the chance to be a student leader and help build something that will outlive me for years to come. Jeff, while you find a way to always put me on the defensive, you've been a valuable part of my experience here. You let me conduct my research rotation in the child acquisition lab, advised

my minor area paper, and sat on both my 895 and Dissertation committees. Your input has been invaluable and in large part responsible for the shift in my interests that has lead to the research I am leaving UMD to embark upon.

Many others have played a role as well. Tonia and Peggy, thank you for the opportunity to teach with you and for helping mentor me as an instructor in Linguistics. I learned as much from teaching sections and observing the classes as I did in my own classes. This experience lead me to complete the University Teaching and Learning Program, which has made me an infinitely better mentor and instructor that I was before. Of course there are also Rochelle Newman and Kira Gor, who served on my dissertation committee and provided invaluable feedback that helped me not only write a better dissertation but rethink some important assumptions I held related to phonetics, speech perception, and second language learning. Kira, thank you also for giving my mother an opportunity to contribute to your research, the personal encouragement over the years, and assisting the cross-langauge priming study years ago, it was my first mentoring experience in linguistics. Kathi, thanks for your help with endless arrangements for conferences, classes, and travel. Kim, thank you for being a friend and constant supporter over the past 5 years. You were always ready to help professionally and personally. I will always remember your kindness. Csilla, thank you for being an administrator extraordinaire and a good friend to boot. I've enjoyed your company and unique view on the world. I don't know how you balance everything you do...and I still hope I can meet your horse one day.

Many other friends have helped me greatly along the way. Sol, words cannot

explain how much your friendship has meant to me. You helped me understand what it takes to be a researcher and never failed to remind me when I wasn't living up to it. Your matter of fact view of the world and those in it helped me through many tough times, and when the going wasn't as tough, your company helped light up many a trip away or evening in. Mathias, you've been a great friend and I've learned a lot from you. I will miss giving you haircuts in the basement. And it only took us 5 years to get that fricatives paper out the door! Dan, Darryl, I was really glad to have you as part of our cohort. Darryl, you were there before even our first day to help move my parents in, and you never failed to come early to prepare food for a birthday celebration, and always with the best bottle of scotch. I was happy to have you as an officemate, even if I never did learn to appreciate Dr. Who the same way. Dan, it was a fun five years. I wish we could have gotten more frisbee and disc golf in and some more guitar jam sessions, but I'm happy that we've grown our friendship and will continue to do so in the future.

There were many people who helped me with research along the way. Erin Bennet and Emily Coppess, you were amazing RAs and I couldn't have had two better people to work with. Erin, you also became a good friend (and sometimes roommate), and taught me so much, I didn't expect to own 2 dozen juggling balls and a unicycle by the time I finished my PhD. Shannon Barrios and Matt Winn, I will simply thank you for the second half of my dissertation. Matt, your prowess with Pratt and acoustic phonetics amazes me. Shannon, you taught me that R coding can actually be fun...and that there is light on the other end of grad school. So many others at Maryland made my time here enjoyable, productive, and worthwhile.

Thank you to all the members of the IGERT program who have ever participated in an outreach event, you helped build something special together. Thank you to all the RAGNAR runners for making coming back to school after the summer a little extra stressful (but oh so worth it). Thank you to Anton, Ilya, and Natalia for helping remind me that at the end of it all, I'm just a Russian boy at heart.

But even with the best advisors, colleagues, administrators, and academic support, none of it would have been possible without those in my life who helped me get here and then helped me get through it. When I was kicked out of school in 2002 I could not have imagined being where I am today. There are several people who deserve special mention for guiding me to the point of where I was ready for this challenge. Joe Bush and Jamie Gagnon, thank you for showing me that its possible to get a PhD and still have a life. Leticia, thank you for literally seeing me through the process of applying to go back to school and for being a great friend and companion. I cherish your friendship more and more with every passing year. Steve O'Neill, I don't think any of this would have been possible if you didn't take me out for coffee to discuss a crazy idea of starting a non-profit together. EPOCA not only became a symbol of my own transformation, but gave me the self-esteem and focus to rise up and move forward in life. On that note, a special thank you to the entire EPOCA crew, we did so much together. The lessons you have all taught me permeated my graduate career and will continue to shape me as a socially-conscious academic for the rest of my life.

Then there are the people who have helped me in ways that are almost impossible to put into words. Tara, thank you for reminding me of where I came from

and for helping one of my life dreams come true. Anna, thank you for reminding me just how big the world is and just how small it can feel. I never expected to have adventures like we did, you literally helped change my view on the world. Sarita, you helped me discover an important part of myself, which I otherwise might now have for many years to come. Firefly, thanks for helping me lift the cover and take a better look at myself and those around me. Em, thank you for being there with me, I'm sorry that I couldn't always be there for you. Abbe, whether in the MMH hallway, in a basement in Silver Spring, or in DC, you've been a great friend and so much more. Its been so rewarding to watch where you've gone over the years and share my time with you. Sarah, you saw me at my worst and still accepted me as I am, perhaps you saw me as better than I am. Thank you for letting me share my life with you, and for sharing yours with me. I literally could not have made it through the final days of the dissertation without you. Thank you for that, and for so much more.

As I wind down this Oscar-worthy acknowledgements section, there are so many others that I need to thank and so many that I am sure I'm forgetting. Piotr and Senthil, thank you for being great roommates and friends, and for helping to take care of Click on my numerous times out of town. I'm glad she found a friend in the two of you, especially you Piotr...I knew you have a special place in her heart, and vice versa. On that note, thank you Click. As cheesy as it sounds, you have been the most stable companion and bedrock, seeing me through getting kicked out of school, jail, running for office, starting my noon-profit, starting my first business, and finally returning to academia. You're a rock!!! Thanks to all the guys and gals

at the New Deal, a home away from home where I got to explore my creative side and share a part of me that lay dormant for a while. And thank you to my soccer team in DC. We had 14 great seasons together, and you made Sundays more than just a day to work without being interrupted. For that I thank you. And I forgive you guys for contributing to my knee surgery (2 weeks before dissertation was due). On that note, thank you for everybody who helped me get through the recovery and made it possible to put the finishing touches on the dissertation. Verb, Pete, Stef, Sirri, Anita, thank you.

Lest I forget, I also want to thank, in no specific order, and for no particular reason (ok, some particular reasons), the following people. You all know what you did. Corey and Jamie, literally unforgettable. Dan and Courtney, you guys are irreplaceable. Amy Baxter, thanks for the impromptu support sessions and the excuse to dress like a lumberjack. Linda, thanks for opening your home to me. Geoff and Giselle, thanks for the quiet evenings out of the city and the reminder that political and religious ideologies can't ruin a friendship. Eric, Alyssa, Lizzy, thanks for keeping Rhode Island feeling like home. Chris, thanks for the New York hangouts and coming down for those long runs. Pat Byrnes, we've come a long way buddy! Lidia, I'm honored to have been a part of the journey.

On a more formal note, I would like to thank the UMD community and administration. For their financial support, but also for the many opportunities to get involved and give back to the academic community. Dean Caramello, it was a pleasure serving on the Graduate Assistant Advisory Council with you and showing that graduate students and administration officials can work together to make

something meaningful happen. Dean Thornton-Dill, thank you for the chance to help bring people together in ARHU and build a more cohesive college. To the University Senate, it was great to come together with undergraduates, faculty, and staff to help guide the university into the future and particularly to pass some important legislation such as the Good Samaritan policy. To the entire Graduate Student Government crew, it was a pleasure to work together on so many issues that matter to all graduate students here. To all my colleagues on GAAC, it was an exciting two years, please keep up the great work. And to all those still struggling and working to build a successful Graduate Student Union, I will always stand with you in solidarity.

Finally, I want to come back to those who have really made it all possible and have supported me through everything and always. That is my whole family. My brothers and sisters. My grandparents. My nieces and nephews. And everybody else. They have known me the longest and still want to spend time with me. That's pretty amazing. And when it came to the big day, they all joined me via youtube (`https://www.youtube.com/watch?v=eYFPldTfSwQ`) to watch me defend my dissertation, being there with me as I completed one of my life dreams and goals. Olga, you hold a special place among everybody...the amount of emotion you pour into our family and all your relationships is amazing and inspiring. Thanks for always checking in on me. For my family's love and constant support I am always grateful. I am a lucky man. And now, thanks to everybody listed here and to many many others I haven't specifically mentioned, I am also a Doctor!

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ID	Identification
Disc	Discrimination
CP	Categorical Perception
PME	Perceptual Magnet Model
VOT	Voice Onset Time
POA	Place of Articulation
L1	First Language
L2	Second Language
PAM	Perceptual Assimilation Model
NRV	Natural Referent Vowel framework
SLM	Speech Learning Model
SLA	Second Language Acquisition
NE	Native English
NR	Native Russian
NELR	Native English Learners of Russian

List of Symbols

τ	Tau, the category:noise variance ratio
d'	D-Prime, a measure of perceptual distance in discrimination studies

Chapter 1: Introduction

1.1 Background Overview

Assigning categories to perceptual input allows people to sort the world around them into a meaningful and interpretable package. This ability to streamline processing applies to various types of input, both linguistic and non-linguistic in nature. Evidence of categorical effects in perception has been found in diverse areas such as color perception[1] (Davidoff, Davies, & Roberson, 1999), facial expressions (Angeli, Davidoff, & Valentine, 2008; Calder, Young, Perrett, Etcoff, & Rowland, 1996), familiar faces (Beale & Keil, 1995), artificial categories of objects (Goldstone, Lippa, & Shiffrin, 2001), speech perception (Liberman, Harris, Hoffman, & Griffith, 1957; Kuhl, 1991), and even emotions (Hess, Adams, & Kleck, 2009; Sauter, LeGuen, & Huan, 2011). Two core tendencies are found across many domains: A sharp shift in the identification function between category centers, and higher rates of discrimination for stimuli from different categories than from a single category. The degree of these effects in particular domains can often vary along a continuum of categoricity, variably affecting different members of the domain.

[1] For color perception, categorical effects arise when the task requires memory recall, encouraging categorical labeling. Without a memory component, perception is much more gradual/linear, sometimes showing no evidence of warping (Roberson, Hanley, & Pak, 2009)

Categorical effects arise when our perception is skewed from a veridical mapping of the input toward category centers (Fig 1.1). Researchers have considered these effects as derived from a decision metric (Massaro, 1987a), effects from a prototype representation of a category (e.g., Kuhl, 1993a), sets of exemplars stored in memory and tagged with category labels (Lacerda, 1995), an emergent property of a connectionist representation (neural networks approach) (Vallabha, McClelland, Pons, Werker, & Amano, 2007; Damper & Harnad, 2000; Guenther & Gjaja, 1996, inter alia), and the influence of innate discontinuities in perception (Pisoni, 1977; Eimas, Siqueland, Jusczyk, & Vigorito, 1971). While the number of approaches to examine these effects has been quite broad over the years, there are several models that have been particularly prominent, especially in the domain of speech perception. These include Categorical Perception (CP) (Liberman et al., 1957) and the Perceptual Magnet Effect (PME) (Kuhl, 1991). While technology has evolved, brain imaging techniques have shone light onto the underlying nature of categorical effects, and advanced computational and statistical techniques have given us many new opportunities to recapitulate categorical effects in new ways, these two models have been consistently used to describe effects and to test for categorical effects in new domains.

1.2 Goals of the Dissertation

The primary goal of this dissertation is to build on this history of research in categorical effects in speech perception and provide new insight into possible

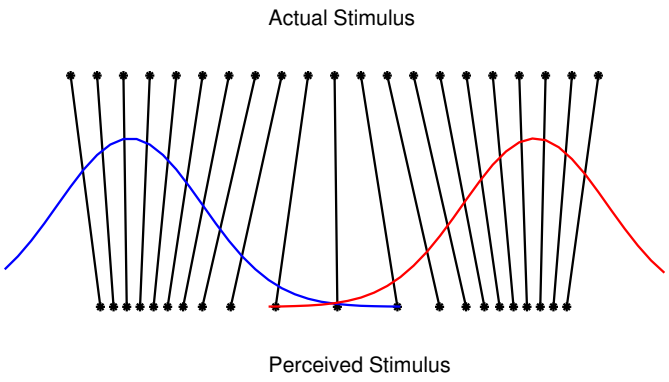


Figure 1.1: This figure illustrates how actual stimuli can be warped toward category centers when they are perceived. Actual stimuli appear on top. The two categories are represented as two distributions, in blue and red. Along the bottom, we see how the stimuli are perceived once categorical effects are taken into account.

underlying mechanisms for these effects. The experiments, simulations, and analyses in this dissertation will be grounded in these classic models. While I acknowledge that neither CP nor the PME are the state of the art in terms of explanatory power and precision of predictions, they do constitute the two behavioral phenomena that have the broadest appeal and deepest history. With this in mind, the first goal of this dissertation is to show that we can derive a computational Bayesian model that can explain the behavioral findings underpinning both effects within a single qualitative framework. Once this is accomplished, the second goal of the dissertation is to apply the Bayesian model to behavioral data from L2 learners to examine how categorical effects change over the course of L2 acquisition. Particularly, I focus on how new categories from the second language coming online together with the first

language categories change the warping in our speech perception.

Because the goal is to examine both classic cases of CP and PME as well as examine foreign language learning, I focus on categorical effects in the domain of speech perception at the level of individual phonemes. First, I will examine the existing findings in categorical effects in phoneme perception and lay out an argument for why these findings can be combined and explained with a single model. Then I will lay out the Bayesian model and show how it can capture these findings in a unified qualitative framework. My goal here is to show how a single parametric quantitative variation can lead to the different degrees of categorical effects seen in perception experiments with different phonemes. I will then examine what happens when we introduce a non-native language and consider behavioral data for listeners who are in the process of acquiring a new language. Specifically, I will gather perception data from native english speakers in the process of learning Russian and use our model to examine the changing nature of category representation with changing L2 proficiency. The goal of this work is to provide a quantitative foundation and the ability to make accurate a priori predictions for identification and discrimination patterns for non-native speech contrasts. Below I lay out in detail how this dissertation is structured and the contents of each section.

1.3 Outline of the Dissertation

In this section I provide a brief overview of the contents of each chapter of the dissertation. I include minimal background information, cover the key topics

covered, and for the experimental chapters provide a quick summary of results.

Overview of Chapter 2: Introduction to Categorical Effects in Speech Perception

In Chapter 2, I review the existing literature on categorical effects in perception and discuss the models of these effects, focusing on their similarities and differences and potential for a unified approach. The models considered here concern the effects of individual native language phonetic categories on perception along a well-defined continuum. Once findings and models for first language perception are laid out, I consider the common models of second language (L2) perception and the notions of similarity and categoricity that are appealed to in the L2 literature. Finally, I consider our ability to unify L1 effects in a single quantitative model and how this model can be applied to L2 learning to make accurate predictions for both naive and eventual L2 perception abilities.

While native language categorical effects have been found for a wide range of phonemes, different phoneme classes differ in the degree to which they are perceived categorically. We can consider the range of categorical effects along a spectrum of how much a role category labels play and degree of sensitivity to acoustic detail. At one end of the spectrum, perception of stop consonants is strongly categorical. Discrimination is little better than would be expected if listeners used only category labels to distinguish sounds, and between-category differences are extremely pronounced (Liberman et al., 1957; Wood, 1976). At the other end of the spectrum,

vowel perception is much more continuous, with some researchers even arguing that vowels display no categorical effects at all (D. B. Fry, Abramson, Eimas, & Liberman, 1962). Researchers have used various mechanisms underlying speech perception to explain these differences. For example, these proposed differences have been claimed to stem from the way each type of sound is stored in memory (Pisoni, 1973) and to be related to innate auditory discontinuities, which seem to influence stop consonant perception (Pisoni, 1977; Eimas et al., 1971). However, despite the difference in the degree of effects or potential processing differences between stop consonants and vowels, the general qualitative trends are very similar. The identification function shifts sharply between the two category labels at the category boundary, typically labeled as the S-curve. The ability to discriminate equally spaced stimuli is increased at the category boundary and is reduced near category centers. These qualitative similarities between the categorical effects in consonants and vowels suggest that these two cases may emerge as instantiations of the same phenomenon. Perceptual differences among different classes of sounds may be purely quantitative rather than qualitative. This raises the question of whether it is necessary to propose two divergent underlying effects to explain the types of categorical effects seen in different phonetic categories.

Phonetic categories are known to play a prominent role in perception for both first language (L1) learners as well as adult learners of a second language (L2). However, the two areas have differed greatly in terms of the presentation of the findings, ranging from intuitive qualitative results to experimentally explicit quantitative results. As of yet, there is little crossover between the well-understood and

explicit models of categorical effects in native language perception and models that consider the role of categories in both native language (L1) and the acquired target language (L2, or TL) in second language acquisition (SLA). In general, we know that phonetic categories influence second language perception and acquisition. One prominent model predicts how a naive listener of a second language will be able to discriminate contrasts based on how these sounds map onto phonetic categories in their L1 (Best, 1994). In Best's Perceptual Assimilation Model (PAM), a naive listener is one who is not actively learning or using an L2, and linguistically naive of the target language test stimuli. In the spirit of contrastive analysis, the model is able to fairly accurately predict difficulties in discrimination, based on the relation of identification patterns of L2 contrasts to L1 categories. Another model attempts to account for the ability, the rate, and eventual level of attainment of non-native phonetic categories (J. E. Flege, 1995, inter alia). In Flege's Speech Learning Model (SLM), the learnability of a new phonetic category is determined by its mapping to L1 categories and based on the assumption that an L2 learner uses the L1 as a template (or filter) and then applies similar learning techniques as does a child. Unlike the models of native language categorical effects, these SLA models do not explicitly quantify distance metrics between individual sounds and existing phonetic categories in terms of perceptual distance, nor do they make concrete predictions as to precise expected performance on any given task.

Overview of Chapter 3: Bayesian Model of Perception and L1 simulations

In Chapter 3, I present the Bayesian model that I use to unify the findings in L1 categorical effects. I then go on to show how the model is fit to behavioral data from various classes of phonemes, confirming that results can be viewed as qualitatively the same, differing only in some quantitative measure. I present various ways of simulating behavioral results and present some original research that provides further testable data for other phoneme classes. Finally, I discuss the findings and consider the implications for other levels of analysis and extensions to other work.

I consider categorical effects on speech perception from a computational perspective; I focus on the computation that the brain is solving in speech perception rather than the particular algorithms involved or the neural implementation (Marr, 1982). To this end, I employ and extend a Bayesian model that was originally proposed by N. H. Feldman et al. (2009). This model considers speech perception as the process of optimal inference of the intended target productions. In the model, variance is treated as coming from two sources, intentional categorical variation and articulatory and perceptual noise variance. The relative weight of these variances determines the strength of reliance on the acoustic signal and the category means, setting up the qualitative framework within which we establish varying categorical effects. We call the ratio of the two variances τ, and by varying τ we move from extremely strong categorical perception to almost continuous perception, representative of the observed behavior described in Chapter 2.

Using this model, we can consider categorical effects in phonetic continua independently of the physical underlying cues. We consider the model as it captures behavioral findings for stop consonants varying along the voice onset time (VOT) continuum, vowels varying in 2-dimensional first and second formant space (F1-F2), sibilant fricatives varying in their central frication frequency, and nasals varying in their initial F2 transitions representing the place of articulation (POA). The vowel simulations are taken from previous work by N. H. Feldman et al. (2009). The stop consonant and nasal simulations are based on data gathered in previous studies by Wood (1976) and Miller and Eimas (1977), respectively. Finally, an experiment was conducted to collect fricative behavioral data and is described in Chapter 3. I show that the model is adequate for all behavioral sets and that the relevant τ ratios fall in an appropriate place on the continuum. Together, the simulations suggest that we can treat categorical effects as arising from a common source independent of the cues employed in perception.

Overview of Chapter 4: Model L2 Application

In Chapter 4, I consider the use of the preceding computational model to explore categorical effects in second language (L2) perception. First, I consider two existing models of second language perception, focusing on naive listeners as well as learners and their eventual attainment. Then I examine the shortcomings in the existing models and how they can be overcome by considering the model from Chapter 3 to examine particular effects. Then I describe a set of experiments examining

phoneme perception by native English second language learners of Russian. Using the results of these experiments, I fit the model to the behavioral data and examine the results to see what they tell us about the category structure of categories in the L2 and how they change as proficiency increases.

The models I consider further are the Perceptual Assimilation Model (PAM) (Best, 1994) and the Speech Learning Model (SLM) (J. E. Flege, 1995) described in Chapter 2. Particularly, I examine their use (or lack thereof) of similarity metrics in making predictions for discrimination of sound contrasts. Upon examination of the models, what predictions they make, and their basis for making these predictions, I expose several critical missing links. First, predictions made by PAM are very coarse, both in the binning of possible assimilation patterns as well as only examining discrimination of prototypical L2 contrasts. If we want to get a better sense of categories' effect on perception, we need to examine stimuli along a continuum between these L2 phonemes. Second, both PAM and SLM predictions are based on behavioral data by participants in an experiment. There is no independent relationship between categories in the language and expected discrimination and new category formation. An ideal comprehensive formalization of categorical effects in L2 would allow us to examine predictions for both identification and discrimination based purely on underlying category representations. Critically, we need a formal way to compute a similarity measure based on the inferred target productions. Finally, we need the ability to use a model to track perceptual category representation over time (as opposed to relying on production data). I go on to show how the model can be extended to multiple dimensions (building off work

presented in Barrios (2013)) and how it can be used to address these shortcomings.

I then describe the series of experiments conducted on L2 learners of Russian that form the basis of my investigation of second language categorical effects using our Bayesian model. An identification, a discrimination, and a goodness judgment study were completed examining vowel perception in the triangular two-dimensional continuum covering the high non-rounded vowels. The endpoints of this continuum were the vowels /i/, /ɪ/, and /ɨ/. Unlike many previous experiments on L2 perception, the identification experiment allowed participants to select categories from the learned second language and not just the native language categories. Further, instead of only examining discrimination between the endpoint stimuli, I examined discrimination of pairs all along the two-dimensional continuum. This allows us to examine predictions for discrimination based on assimilation patterns at a much finer-grained level. These finer grained measures also allow us to examine how these assimilation and discrimination patters relate to the distance between the stimuli along the F1/F2 continuum as well as the predicted perceptual distance based on perceptual warping due to the categories involved. I also describe the novel approach used for stimulus creation for the experiment, which allowed us to avoid some common pitfalls regarding diphthongization.

Finally, I describe the simulations used to fit the model to the behavioral identification and discrimination data collected from the L2 learners. This allows me to examine the most likely values for prior probabilities, category means, category variances, and noise variances for learners at different levels of proficiency as well as native speakers of both English and Russian. With these simulations, we can

examine which of the properties of the category representation are being set early on and which change with proficiency, eventually leading to accurate perception. Critically, we can get a sense of the level of proficiency needed for the category representation of L2 vowel categories to reach a stage such that the category begins to exhibit its own effects on the surrounding perceptual space. I also discuss how the data collected relates to our understanding of the predictions made by the PAM and SLM for a continuum of learners rather than just naive listeners or eventual attainment.

Overview of Chapter 5: Discussion and Conclusion

I conclude the dissertation with Chapter 5, where I summarize the experiments and simulations and discuss the key findings and implications of the dissertation. First, I provide a review of the particular findings in the experiments and model fitting discussed in the dissertation, along with relevant parameter estimates. Then, I synthesize the findings from the L1 perception research and show how the Bayesian model accounts for the full range of results. I then synthesize the L2 findings, focusing on the key take-aways and consider what we actually learned about perception across proficiency levels. Then I discuss the deeper theoretical implications of the dissertation and what it tells us about the relationship between categories and perception and L1 and L2 processing. I go on to openly consider some limitations of the model and the present work and put forward some fruitful directions for future investigation. Finally, I conclude the dissertation with the key insights and

accomplishments laid out in the preceding chapters.

Chapter 2: Introduction to Categorical Effects in Speech Perception

In this chapter I review existing studies on categorical effects in L1 and L2 speech perception, consider the relationship of the two fields to each other, and propose a possible unification of L1 results with extensions to L2 learning. The typical approach to examining categorical effects in L1 and L2 research is rather different. In L1 research, categorical effects are usually examined from the point of view of how strongly category centers attract intermediate stimuli and the resulting perceptual warping in the phonetic space (Kuhl, 1991; Liberman et al., 1957, inter alia). The focus is not on any consequence for the listener, but rather a theoretical question about the process of speech perception in the brain. In second language acquisition (SLA) literature, the focus on categorical effects is on how the categories in the native language affect perception of individual contrasts in the L2 and also how the alignment of categories in the two languages affects development and eventual attainment (J. E. Flege, 1995; Best, 1994, inter alia). In SLA literature, the focus does not tend to be on perception of a continuum of stimuli in the phonetic space nor on how the categories of the two languages work together to warp perception. Nor is the focus usually on the mental representation of the speech categories, but rather on the production and perception of speech, along with an underlying question of how

to improve the second language process or increase nativeness (F. E. Flege, Bohn, & Jang, 1997; Iverson, Hazan, & Bannister, 2005, inter alia) (though these training studies are a rather different breed from the naive perception and attainment studies which tend to be more theoretical). However, despite these differences, the idea that the categories we learn as children influences how we perceive speech sounds as adults permeates both fields. Hence, a model that explains the source of categorical effects in a native language should be able to shine light on cross-language perception and allow us to examine how categories interact and change over time. In this chapter I review the literature from both traditions in order to set the foundation for the modeling work in Chapters 3 and 4.

2.1 L1 Effects

While categorical effects have been shown in a wide range of perceptual domains (Harnad, 1987), two key models of categorical perception (CP) and the perceptual magnet effect (PME) were originally proposed for stop consonants (Liberman et al., 1957) and vowels (Kuhl, 1991), respectively. In this section, I review these initial models and look at their application to other domains and proposed related implementational models.

2.1.1 Behavioral Measures and Perceptual Warping

The measures that we consider in evaluating the effects of categories on perception are identification, discrimination, and goodness. I'll briefly discuss the goodness

task below, but for now will focus on identification and discrimination, which are central to the category fitting work in my dissertation. Identification and discrimination are the two classic behavioral measures that provide insight into the listener's ability to classify the sounds using labels based on categories available in their language (identification) and to differentiate sounds located near each other along some acoustic continuum (discrimination). In both cases, all the studies that we consider concern a scenario where stimuli are presented along a one-dimensional continuum between two prototypical phonemes. For presentation purposes, consider a continuum between two phonemes, labeled c_1 and c_2, with seven stimuli, $s1...s7$, equally spaced along the linear acoustic continuum (e.g. c1=/b/, c2=/p/, stimuli created by varying the voice onset time (VOT) of the signal).

In this scenario, the identification task consists of choosing between two competing labels, c_1 and c_2. This can be done by presenting stimuli in isolation or in the presence of a competing sound. Independent of this difference, all experiments we consider use the "forced choice" paradigm for identification. This means that the participant has to choose one of the two labels for every stimulus heard even if they are unsure of the proper classification. By examining the rate at which participants choose one category over another we learn both the rate of switching between the two categories as well as the boundary between the categories. The shape of the identification curve between the endpoints provides information about the distribution of sounds in the categories that the listener expects to hear. If perception is strongly categorical then we would expect more absolute identification and a fast switch between category labels. However, a sharp identification curve could also be

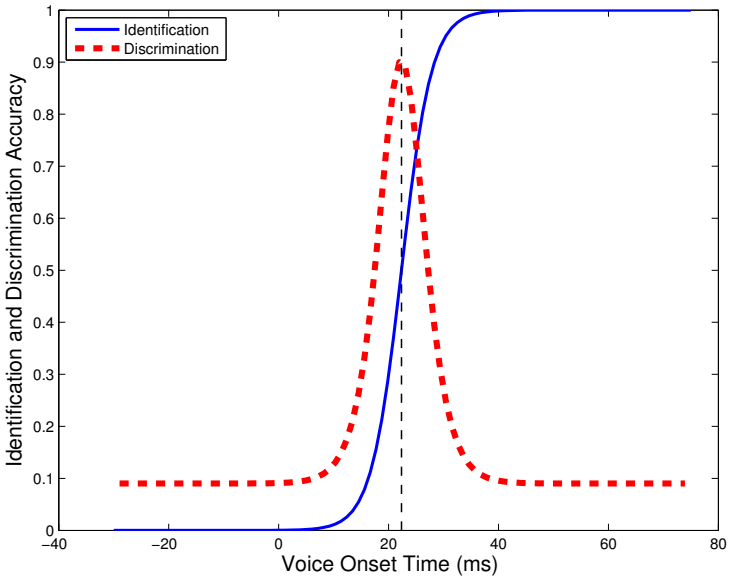


Figure 2.1: Identification and Discrimination in the presence of strong category influences.

due to both categories having very little variance and hence no overlap between their distributions. In this situation, even if perception is uninfluenced by categories, and is strictly gradual, the forced-choice identification curve would still be sharp. The sharpness of the curve then serves as a measure of overlap of possibility of belonging to either category, whether due to perceptual warping toward the categories or the category variances. Later in the dissertation I will discuss how these are related in the Bayesian model I consider. An illustrative example of a sharp identification curve, in this case due to strong categorical effects, is presented in the solid identification line in Figure 2.1.

Discrimination can be done in a number of ways (e.g. AX, ABX, Triad presentation), but all of them test the ability of a listener to tell whether sounds are the same or different. The simulations presented here employ studies utilizing the AX

discrimination paradigm (vowels, fricatives, stop consonants) and the triad model (nasal consonants). The AX discrimination paradigm is one where listeners are presented with two stimuli, which we call A and X. Their task is to say whether the X stimulus is identical to the A stimulus or different from it. The triad model presents a set of three stimuli where two are the same and one is different and the participant is asked to identify the odd man out. For every pair of stimuli A and B there are six possible triads that are presented, including AAB, ABA, BAA, BBA, BAB, and ABB. For these six triads, the correct answer is 3, 2, 1, 3, 2, and 1, respectively. Both tasks measure how likely discrimination is to occur, and therefore serve as a metric of perceptual distance between the stimuli. For a detailed consideration of different discrimination experiment designs from a signal detection theory (SDT) perspective, see Macmillan and Creelman (2005).

By considering equidistant pairs of stimuli along the continuum we can see how listeners' ability to discriminate sounds changes as we move from the category centers to the category boundary found in the identification task. The distance doesn't matter as long as it is kept constant and we avoid floor and ceiling effects. Typically 1-step or 2-step discrimination is examined depending on number of overall stimuli in the continuum. With no warping due to categorical effects, we would expect uniform discrimination along the continuum, due to the equal spacing of the stimuli (e.g. we should be just as good at telling s1/s2 apart as s3/s4). With a stronger influence of the involved categories biasing our perception toward the category centers, our ability to differentiate stimuli near category centers should go down while our ability to tell apart stimuli at the category boundary should go

up. An illustrative example of discrimination performance in the presence of strong influences of categories can be seen in the dotted discrimination line in Figure 2.1. The degree of warping in discrimination is a key indicator of how strong the category effects are. It will be particularly key in our generative Bayesian model for teasing apart different sources of variability in individual category structures.

Finally, there is the goodness task.This task usually employs a Likert scale (typically a scale from 1 to 5 or from 1 to 7) to gauge how well a certain stimulus represents a specific category. Typically, a participant in a study is presented with a particular category, possibly with some training or familiarization, and then asked to rate a list of stimuli as to how well they represent this category. Goodness ratings can provide an approximation of the size of the category by looking at how fast they degrade as we move away from the category center. Additionally, goodness ratings serve as a good measure of within-category variability by observing how the goodness ratings vary for stimuli that are classified as the same category in the identification task. In L1 research, goodness ratings are sometimes correlated with discriminability to see if they predict where discrimination will be better or worse (see the perceptual assimilation model in Section 2.1.3). In L2 research, goodness ratings are used to tell apart assimilation patterns that put two sounds into the same category as equal representatives vs. sounds that are identified as the same category but with varying reliability (see the perceptual assimilation model in Section 2.2.1). While in this dissertation I never model goodness ratings directly, they will come up in various discussion concerning model fitting in both the L1 and L2 chapters.

2.1.2 Stop Consonants and CP

Stop consonants have consistently been found to exhibit strong effects of categories in behavioral tests of perception. Liberman et al. (1957) showed that discrimination of stop consonants is only slightly better than would be predicted if listeners only used category labels produced during identification and ignored all acoustic detail. They labeled this observation "categorical perception". The core tenet of pure categorical perception (CP) for stop consonants is that participants do not pay attention to small differences in the stimuli, rather treating them as coarse categories, ignoring some of the finer detail in the acoustic stream. Under the CP hypothesis, participants assign a category label to each stimulus and then make their discrimination judgment based on a comparison of these labels. This behavior produces the strongest possible categorical effect. Liberman et al. (1957) considered the place of articulation continuum from /b/ to /d/ to /g/ in an artificial CV pair with /e/. They had the participants perform identification of the stimuli in isolation and a forced choice ABX discrimination task. In the identification task, they found a sharp change in the identification function near category boundaries for each of the phonemes. In the discrimination task, they found a peak in discrimination at the same location at which they found the sharp changes in the identification function.

Liberman et al. formulated a probabilistic model which used the probabilities from the identification task to make predictions about how often the listener would be able to discriminate the sounds based only on category assignments. They found that using this formula they could predict the overall discrimination behavior very

well using only the pure identification information. In fact, the participants' actual discrimination only out-performed the predictive model slightly. However, the fact that the participants did outperform the model suggests that they were able to use some acoustic cues beyond pure category membership.

After the initial findings by Liberman et al., many researchers investigated other phonetic environments to see where else similar categorical effects appeared and how these categorical effects are modulated by anything from contextual effects to task-related factors (Pisoni, 1975; Repp, Healy, & Crowder, 1979). Critically for our work here, findings similar in nature to those of Liberman et al. for place of articulation were found by Wood (1976) for the voicing dimension (ie: the voice onset time (VOT) continuum). Categorical effects such as these have also been found for /b/-/d/-/g/ by other researchers (Eimas, 1963; Griffith, 1958; Studdert-Kennedy, Liberman, & Stevens, 1963, 1964), as well as for /d/-/t/ (Liberman, Harris, Kinney, & Lane, 1961), /b/-/p/ in intervocalic position (Liberman et al., 1961), and the presence or absence of /p/ in 'slit' vs. 'split' (Bastian, Delattre, & Liberman, 1959; Bastian, Eimas, & Liberman, 1961; Harris, Bastian, & Liberman, 1961). Because of this wide range of research, CP is generally accepted as a robust effect by researchers in the field.

Various models have been put forward to explain the source of categorical perception. Initially, researchers assumed that categorical perception resulted from psychophysical properties of processing speech and argued that it was specific to language processing (Macmillan, Kaplan, & Creelman, 1977). Other researchers tended to use more general views of either statistical properties or higher order cognitive pro-

cessing to explain the effect. Massaro (1987a) used signal detection theory (SDT) to model the identification and discrimination tasks in two stages, sensory and decision operations, leading to a separation of sensitivity and response bias. In that model, categorical behavior arise from classification behavior even if perception is continuous, with Massaro calling it *categorical partition* instead of *categorical perception*. The separation of perception and decision making processes was further investigated by Treisman, Faulkner, Naish, and Rosner (1995), who applied criterion-setting theory (CST) (Treisman & Williams, 1984) to categorical perception. Their work models the sensory system as able to reset the internal criterion for decision making based on most recently available data, much like Bayesian belief updating. Elman (1979) showed that such a model of criterion setting is able to capture the original stop consonant findings (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967) even better than their original Haskins model. Other researchers considered the problem at a different level of analysis, focusing instead on the possible neural implementation of the categorical perception mechanism. Vallabha et al. (2007) proposed a multi-layer connectionist model that operates on Gaussian distributions of speech sounds as the input and produces categorical effects via interactions of three levels of representation: an acoustic input layer, an intermediate perceptual layer, and a categorical classification output layer. The key to their model is the presence of bidirectional connections between the output category level and hidden perceptual layer, whereby the perception influences the classification, but the classification simultaneously biases perception toward category centers. The setup of the model and use of bidirectional links to create top-down influences is similar to the TRACE

model of speech perception proposed by McClelland and Elman (1986), where a feature level, phoneme level, and word level were used to explain various features of speech perception. Other neural network models focused on lower level neural activation patterns as a mechanism by which these categorical perception effect could arise. Damper and Harnad (2000) trained both a Brain-State-In-A-Box (BSB) (following Anderson, Silverstein, Ritz, and Jones (1977)) and a back-propogation neural network model to show how categorical perception arises through spontaneous generation after training on two endpoint stimuli. They were able to produce typical categorical effects and reproduce the discrepancy between VOT boundaries between different places of articulation found in human participants. Focusing on individual neuron activations, Salminen, Titinen, and May (2009) exposed a self-organizing neural network to statistical distributions of speech sounds represented by neural activity patterns. Their resulting neural map showed strongly categorical effects from single neurons being maximally activated by prototypical speech sounds, along with the greatest degree of variability in the produced signal at the category boundaries. Kröger, Birkholz, Kannampuzha, and Neuschaefer-Rube (2007) showed that categorical perception arises when using distributions consisting of specific features (bilabial, coronal, dorsal) to train self-organizing maps to learn phonetic categories and discriminate between sounds. These models suggest that there are many possible processes that could underlie strong categorical perception as originally described by Liberman et al. and many others. However, these models are poorly adapted to capture effects going beyond the case of strong categorical perception described above.

While both the behavioral and modeling data do make a case for strong effects of categories, we have accumulated evidence that shows us that perception of stop consonants is not purely categorical. Listeners pay attention to sub-phonemic detail, as evidenced by various behavioral and neural studies. Studies have shown that goodness ratings of stop consonants vary within categories and are prone to context effects based both on phonetic environment and speech rate (Miller, 1994). Internal structure of consonant categories is further supported by studies of reaction time. Pisoni and Tash (1974) showed that "Same" responses were faster to identical rather than different within-category stimuli. Further, they showed that answers of "Different" to stimuli on the category boundary with large acoustic differences was faster than boundary-spanning stimuli with low acoustic diffrences. Together, suggesting that listeners had access to some fine acoustic detail in stop consonant perception. Further, priming studies by Andruski, Blumstein, and Burton (1994) found priming effects of within-category VOT differences for short inter stimulus intervals of 50 ms. They showed that stimuli with initial stop consonant VOTs near the category center exhibited a stronger priming effect for semantically-related following stimuli, and that non-central values also elicited longer reaction times. Finally, at the neural level, an fMRI study by Blumstein, Myers, and Rissman (2005) showed that there are robust neural correlates to sub-phonemic VOT differences in stimuli. Together, these studies strongly suggest that not all members of the category are truly equal and that identification and discrimination performance cannot be based on an all-or-none scheme.

2.1.3 Vowel Perception and PME

Vowels exhibit less influence from categories and tend to be perceived more continuously, with listeners exhibiting higher sensitivity to fine acoustic detail. Their discriminability cannot be predicted purely from the identification data, which itself is much more prone to context effects (Eimas, 1963). This has been found for the /i/-/ɪ/-/ɛ/ continuum (D. B. Fry et al., 1962; K. N. Stevens, Ohman, & Liberman, 1963; K. N. Stevens, Ohman, Studdert-Kennedy, & Liberman, 1964). Additionally, the same core finding was observed for perception of vowel duration by Bastian and Abramson (1962) and for tones in Thai by Abramson (1961). These findings led researchers to claim that vowel perception behaved much like non-speech signals, with no effect of phonetic categories on listeners' abilities to encode and decode them. Further support for more continuous perception of vowels came from mimicry experiments (Chistovich, 1960; Kozhevnikov & Chistovich, 1965). When participants were asked to mimic stop consonants and vowels, their ability to reproduce vowels accurately was much greater than for consonants, which tended to be reproduced with prototypical members of the category. It should be noted that these findings are all for steady-state vowels and does not address vowels in speech contexts that contain rapidly changing formant structures. K. N. Stevens (1966) found that one could obtain nearly categorical perception when looking at vowels between consonants pulled out of a rapidly articulated stream. However, for comparisons in this work, I focus on findings in the standalone vowel case.

A different approach to investigating the role of categories in vowel perception

came when Kuhl (1991) conducted experiments considering what sorts of categorical effects appeared in identification, discrimination, and goodness ratings of various vowels along a continuum. In this context, goodness ratings are ratings on a fixed scale of how well a particular stimulus represents a specific category. In her study, she collected goodness ratings for stimuli equally spaced around a particular vowel that was either a prototype of the category or a non-prototype stimulus. The choice of the prototype and non-prototype vowels was based on goodness ratings from Grieser and Kuhl (1989) for a continuum of /i/ vowels from the range originally defined in Peterson and Barney (1952). The collected goodness ratings confirmed that there was variable within-category structure that people could represent and access; multiple participants shared the center of goodness ratings (i.e. the location where stimuli were rated highest on category fit) and had similar patterns of judgment for stimuli expanding radially from the center. These findings suggest that participants have a stable representation of the category for /i/. Further, they suggest that the structure does not represent an all-or-nothing judgment of category membership for all tokens, but rather a gradient representation.

The next question was how this gradient representation relates to the perception and discriminability of individual stimuli. To answer this, adults, children, and monkeys were asked to discriminate sounds equally spaced around the prototype and non-prototype centers along a linear continuum connecting the centers shown in the goodness rating study, P (Prototype) and NP (Non-Prototype). Both adults and children were more likely to perceive stimuli around the prototypical category member as the same sound as compared to sounds around the non-prototype. However,

monkeys did not show any effect of prototype or non-prototype. This suggested that humans of both ages were using linguistically-informed representations of category structure to guide their interpretation of similarity of sounds to each other.

These findings led Kuhl to propose the perceptual magnet effect (PME). Specifically, she described this effect as a within-category phenomenon, focusing on the relationship between category goodness judgments and ability to discriminate neighboring vowels. The claim was that stimuli that are judged to be better exemplars of a category act as "perceptual magnets", pulling surrounding stimuli together, thereby making them harder to discriminate. Meanwhile, stimuli judged to be poor exemplars exhibit very little effect on neighboring vowels. As a result, under the perceptual magnet hypothesis, there is a correlation between category goodness judgments and discriminability. Critically, the effect does find something like categorical effects, with additional effects from the goodness for the stimuli. This was further investigated using signal detection theory and a multi-dimensional scaling approach in Iverson and Kuhl (1995). They showed that there is a direct link between goodness of stimuli and their discriminability, proposing that this is a central tenet of the PME. These findings showed that vowels are not in fact continuously perceived, though the precise nature of the effect differs from that of stop consonants.

The finding of the PME for vowels is a critical one for the present work, as the Bayesian model that I adapt for my simulations was designed to investigate these PME findings in order to see if an optimal inference mechanism can be used to show where such effects might be coming from (N. H. Feldman et al., 2009). Additionally,

Kuhl observed that her proposal might account for findings in perception of foreign languages. Particularly, researchers have shown that people assimilate phonemes in a foreign language to close prototypes in their native language (Best, McRoberts, & Sithole, 1988; J. E. Flege, 1987). Additionally, we know that children, by 10 to 12 months of age, begin to fail to discriminate sounds in a foreign language in pairs where the distinction is not linguistically informative in their native language (Werker, Gilbert, Humphrey, & Tees, 1981). This suggests that the representations that led to the perceptual magnet effect may be online very early in linguistic development.

The extent to which the PME generalizes to other sound types is an open question. There were documented reproductions of the PME for the /i/ category in German (Diesch, Iverson, Kettermann, & Siebert, 1999) and Swedish (Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Aaltonen, Eerola, Hellström, Uusipaikka, & Lang, 1997), but also failed replication attempts for American English (Lively & Pisoni, 1997; J. E. Sussman & Gekas, 1997) and Australian English (Thyer, Hickson, & Dodd, 2000). Additionally, there was a failure to find evidence of the PME for certain other vowel categories in English (Thyer et al., 2000). However, it was also found for the Swedish /y/ category (Kuhl et al., 1992; Aaltonen et al., 1997) as well as the lateral and retroflex liquids (/l/,/r/) in American English (Iverson & Kuhl, 1996; Iverson et al., 2003). These findings for various vowels suggest that the effect is capturing meaningful patterns in perception, even it the precise nature of the stimuli that elicit it is unclear.

Several models at different levels of explanation have been proposed to explain

the source of the PME. One such model is the Native Language Magnet Theory (Kuhl, 1993a). This was the original theory put forth by Kuhl to show how prototypes of categories affected our perception and lead to the PME. However, this model did not address the question of why prototypes should exert a pull on neighboring speech sounds. An exemplar model was then proposed (Lacerda, 1995) that showed how the PME could be construed as an emergent property of an exemplar-based model of phonetic memory. In Lacerda's model, sound perception is guided by a simple similarity metric that operates on collections of exemplars stored in memory, with no need to refer to special prototypes to derive the sorts of effects typical of the PME. This then left the question of how do we fully account for within-category discrimination. For this we can consider low-level neural network models that attempt to provide an explanation of the type of connectionist network that can give rise to these perceptual effects. One such neural network model was proposed by Guenther and Gjaja (1996), where sensory experience guided the development of an auditory perceptual neural map and the vector representing cell firing corresponded to the perceived stimulus. In another model, Vallabha and McClelland (2007) considered learning via distributions of speech sounds using online mixture estimation and Hebbian learning to derive the effect. Both models showed how the effect might be derived from a biologically plausible mechanism.

2.1.4 Fricatives

Fricatives are interesting to consider in relation to the other presented work since they share properties with both the stop consonants and vowels. They are a type of consonant; however, they also share properties with vowels, in that they can largely be identified and discriminated by their spectral properties. Particularly, by the spectral peaks in the frication noise (Jongman, Wayland, & Wong, 2000). The precise spectral frequency cues are different in that fricatives have higher frequency aperiodic noise and vowels consist primarily of lower frequency periodic energy. However, they are qualitatively similar, in that this spectral information is the key to the identification of the particular sounds. Some consonants also employ spectral cues like F2 onsets for place of articulation, but that is different from a steady state frequency band. Because of this similarity, fricatives serve as an interesting case to explore perception behavior that may fall intermediate between vowels and consonants.

Attempts to examine categorical perception in fricatives have had mixed results, with divergent findings between researchers, and sometimes even within the same research project. In experiments with fricatives, Repp (1981) found that in behavioral trials participants' behavior was similar to that originally found in trials with stop consonants, indicating categorical perception of fricatives. However, during the course of the same study, some of the participants in the experiment exhibited perception behavior that was much more continuous. To accommodate this apparent contradiction in the findings, Repp proposed that participants were using

two distinct processing strategies: acoustic and phonetic processing. Phonetic processing refers to a mode of perception where listeners are actively assigning phonetic category classifications, whereas acoustic processing refers to a mode of perception where listeners are attuned to the fine-grained acoustic variability of the signal.

Other researchers, either via experiments on identification and discrimination or by considering the types of cues related to perception, have made claims that fricatives do not exhibit categorical properties like stop consonants. In their work on classifying the properties of perception of various forms of speech, Liberman et al. (1967) considered a sound's tendency to show restructuring, or exhibiting varying acoustic representations as a result of varying context, and how this related to observed categorical effects. Stop consonants were found to exhibit a large amount of restructuring, or changes in how they appear acoustically even though they have the same underlying phonemic status. This was found in both correlates of place of articulation (Liberman et al., 1967) as well as manner and voicing (Lisker & Abramson, 1964b; Liberman, Delattre, Cooper, & Gerstman, 1954). Steady state vowels, on the other hand, show basically no such restructuring, when accounting for speaker normalization and speaking rates (Liberman et al., 1967). In considering the noises produced at the point of constriction in both fricatives and stop consonants, Liberman et al. claim that for longer duration of the noise, precisely the kind in fricatives, the cue does not change with context, hence no ability to produce restructuring of the stream perception. This was shown specifically for the perception of /s/ and /ʃ/ by Harris (1958) and Hughes and Halle (1956). This particular metric of restructuring suggests that fricatives should pattern with vowels and not

stop consonants in terms of categorical effects. This finding stands in opposition to the main finding from Repp (1981), where most participants exhibited strong categorical effects.

Following up on previous inconclusive work, Lago et al. (2010) investigated categorical effects in fricatives by focusing on the continuum between sibilant fricatives /s/ and /ʃ/. They conducted an identification task, an AX discrimination task, and a goodness judgment task. Their results showed a strong effect of categories on the perception of the stimuli with no strong correlation between discriminability and goodness ratings. Qualitatively, their identification findings showed the expected S-curve with a sharp change in identification near the category boundary, but a discrimination peak that was markedly shallower than that typically found in stop consonant experiments. This suggested that the fricatives employed a representation that incorporates more of the acoustic signal beyond pure category assignment, making them not as strongly categorical as stop consonants, though also not as continuous as vowels.

2.1.5 Nasals

Nasal consonants constitute the final phoneme category I consider in this work. Nasal consonants are acoustically similar to oral stop consonants, but with a nasal resonance (also called nasal murmur or major nasal formant) that precedes the onset of the formants and carries over into the proceeding vowel. Within nasals, when distinguishing the place of articulation, the difference is primarily cued by the formant

transitions for both F2 and F3, just as it is for oral stop consonants in the same place of articulation (Larkey, Wald, & Strange, 1978). If this were the only cue used for identification then we would expect very similar performance to that of oral stop consonants. However, the nasal murmur can also serve as a cue for identification (Mayo & Turk, 2005), raising the possibility of more continuous perception. In additional, nasal stop consonants can also be identified by an antiresonance (or nasal zero) in the spectrum. The acoustic theory of speech perception makes clear predictions for the structure of nasal consonants (Fant, 1960; Fujimura, 1962; Flanagan, 1972). Particularly, the blocked oral cavity acts as a side-branching resonator, introducing valleys in the spectrum, with the frequency of the nasal zero determined by the shape of the side-branching resonator. Hence, the nasal zero moves as we change the place of articulation, moving from /m/ to /n/. While extracting these parameters has proven difficult to implement in machine speech recognition applications it is in theory a reliable cue for human listeners (Yingyong & Fox, 1992).

Beddor and Strange (1982) investigated whether there were language experience effects on perception of the nasal formant by examining the oral-nasal contrast for both consonants and vowels with Enlgish and Hindi speakers. The consonant contrast (/ba/-/ma/) is phonemic in both languages, while the vowel contrast (/ba/-/bã/) is only phonemic in Hindi. What they found was that there were no language differences, with both Hindi and English speakers treating the continuums categorically for both consonants and vowels. However, for the vowel case they found that language differences emerged when the ISI was increased and they extended the range of the velar port opening.

Larkey et al. (1978) also wanted to investigate the effects of language experience on the perception of nasal consonants. However, they were not looking at the transition between a phoneme with and without a nasal formant, but rather the place of articulation contrast for nasal stops in the presence of a nasal formant for all stimuli. They investigated nasal perception as cued only by F2 and F3 onset formant transitions in both syllable-initial (/mae/-/nae/-/ŋae/) and syllable-final (/aem/-/aen/-/aeŋ/) positions. They found that the /n/-/ŋ/ contrast had a discrimination peak syllable-finally but the peak was absent syllable-initially, suggesting that context is relevant to discrimination (/ŋ/ does not occur syllable initially in English). They showed that this was not the case for the oral counterparts of /d/-/g/, indicating that the difference was linked to the presence of the nasal formant. This result showed that perception of certain linguistic dimensions is constrained by language experience and exposure to these contrasts in certain phonetic environments. Data matched categorical perception predictions closely, with discrimination performance being predicted accurately by just the identification performance.

In another study, Miller and Eimas (1977) simultaneously examined perception of place or articulation and manner continua by looking at labial-alveolar and nasal-oral stop distinctions. They had listeners perform identification and discrimination along the /ma/-/na/, /ma/-/ba/, /ba/-/da/, and /na/-/da/ continua. They found evidence of categorical perception along all four continua, with cross-category discrimination performance substantially better than within-category discrimination. These findings suggested that strong categorical perception was something inherent to consonants, since in the nasal-oral continua the cue being changed was a formant

whose spectral properties resembled that of vowels.

Together, the research on nasal consonants suggests that they are prone to strong categorical effects. This was found whether you investigate the continuum between oral and nasal consonants or between two nasal consonants. The only place where strong categorical effects were not found was for the alveolar-velar continuum with nasals in word initial position. However, because the velar nasal does not appear in English in this position, then this is not really a case of a continuum between two valid phonemes and does not act as a counterpoint to the general finding of strong categorical effects.

2.1.6 Common Ground in Vowel and Consonant Perception

A cursory look at the different degree of categorical effects for different phonemes does reveal some substantive differences. Stop consonant perception is characterized by very sharp identification shifts between two categories and a large peak in discrimination at the center between the categories. Vowel perception elicits more continuous identification functions, shallower peaks in discrimination at the category boundaries, and much greater within-category discrimination. Additionally, goodness ratings for vowels follow a gradient descent from the center of the category outward (Iverson & Kuhl, 1995), while for stop consonants the goodness ratings barely vary within the category (particularly for /p/), while they exhibit a sharp jump in goodness at the category boundary (Miller & Volaitis, 1989). Models proposed for these effects only tend to work for one type of category. For stop consonants, the

#	**Categorical Perception**	**Perceptual Magnet Effect**
1	Poor within-category discrimination	Graded within-category discrimination
2	Greater discrimination across category boundaries	Decreased discrimination around category centers
3	Little within-category gradation	Graded goodness judgments for category members
4	No correlation between goodness judgments and ability to discriminate	Correlation of goodness judgments and discriminability

Table 2.1: Categorical Perception vs Perceptual Magnet Effect

Liberman et al. (1967) model can predict discrimination based on the identification function, while this prediction fails in vowel perception experiments. Neural network models tend to only be applicable to vowel perception (Guenther & Gjaja, 1996; Vallabha & McClelland, 2007) or to stop consonant perception (Damper & Harnad, 2000), but the same models have not successfully accounted for perception of both classes. And when we look at other phonemes such as fricatives and nasals, while nasals seem to closely pattern with stop consonants and therefore might be explainable by the same model, fricatives do not neatly fit into either of the categories. It is then understandable that the idea of separate phenomena of CP and PME has persisted with time.

However, perception of the different sound classes has much in common qualitatively. All phonemes examined above exhibit greater discriminability at the category boundaries, with the boundaries matching those found in identification curves. Also, it has been shown that stop consonants do in fact exhibit some within-category discriminability, or at least within-category structure, as evidenced by reaction time measures (Pisoni & Tash, 1974; Massaro, 1987b). It has also been suggested that

the difference in the ability to predict discrimination from identification is due to faulty methods and not an inherent difference in perception. Lotto, Kluender, and Holt (1998) suggest that the effects found in Kuhl (1991) are just another case of greater discrimination of cross-category perception and by retesting identification in paired contexts, they are able to predict discrimination accurately, removing the need to appeal to goodness of stimuli to account for additional variance.

There is further evidence that the various effects may be derivative of a single underlying source. Some experiments suggest that the dichotomy is exaggerated between the two effects. Considering the core tenets of CP and PME presented in Table 2.1, researchers have sought to find contexts fitting one or the other, rarely considering both possibilities. However, when looking at German vowel quality (duration) and considering if their perception exhibited elements of PME or CP, Tomaschek, Truckenbrodt, and Hertrich (2011) found that they exhibited properties of both. Of particular interest, while Iverson and Kuhl (2000) found that goodness of stimuli can be pulled apart from identification and discriminability effects (a key of the PME), Tomaschek et al. (2011) found that they co-occur. Additionally, fricative perception falling between stop consonants and vowels further suggests that the degree of effects may be gradient rather than falling at one end of the continuum or the other, and hence derivative of some parametric variation.

Given the evidence presented, it is worthwhile to pursue a unified theory of categorical effects in phoneme perception. A unified theory makes stronger, more quantifiable predictions instead of relying on local explanations of individual findings. A unified theory also yields new questions and provides a framework for future

investigation. In general, a single explanation is preferred if it captures results typically ascribed to two different phenomena, a kind of Occam's razor. A similar argument was used to substantiate a unified account of cumulative exposure on selective adaptation and phonetic recalibration by Kleinschmidt and Jaeger (2011, 2012). Here, I argue that CP and PME can be considered to be two instantiations of the same phenomenon. First, they exhibit similar behavioral properties, including better discrimination at the boundary, reduced within-category discrimination, and discrimination peaks between categories. Second, even if the perception of the cues to the different sounds is implemented differently at the neural level, the problem of proper speech perception may still be modeled as a single phenomenon at the computational level, in the sense proposed by Marr (1982). Following this logic, I derive a single explanation for the range of categorical effects that we have thus far discussed. First, I present and expand a model that describes speech perception as a process of optimal inference. I then use this model to fit the range of behavioral findings thus far described. Finally, I show how parametric variation within the model leads to gradient categorical effects, concluding that categorical effects in phoneme perception can in fact be unified.

2.2 L2 Effects

Phonetic categories in our native language also influence our ability to perceive contrasts in foreign languages and to acquire new foreign phonetic categories accurately. One prominent model predicts how a naive listener of a second language

will be able to discriminate contrasts based on how these phonemes map onto phonetic categories in their L1 (Best, 1994). In Best's Perceptual Assimilation Model (PAM), a naive listener is one who is "not actively learning or using an L2" and "linguistically naive of the target language test stimuli." Following the spirit of contrastive analysis (Lado, 1957), a failed theory where mappings between the L2 and L1 were used to predict what would be hard to learn in L2 learning, the model is able to fairly accurately predict difficulties in discrimination of sound contrasts in the L2 based on the relationship with classification into categories in the L1. Another model attempts to account for the ability, the rate, and eventual level of attainment of non-native contrasts (J. E. Flege, 1995, inter alia). In Flege's Speech Learning Model (SLM), the learnability of a contrast is determined by its mapping to L1 categories and based on the assumption that an L2 learner uses the L1 as a template (or filter) and then applies similar learning techniques as does a child. Unlike the models of native language categorical effects, these SLA models do not explicitly quantify any of the contrasts in terms of perceptual distance, nor do they make concrete predictions as to precise expected performance on any given task. Rather, they make qualitative predictions for how the L1 categories effect gross performance in the L2. In order to consider in detail how these models relate to my present work, I present some of the history and detailed findings of both models below.

2.2.1 Naive Perception and the Perceptual Assimilation Model

A prominent model of perception of L2 contrasts for the naive leaner is the Perceptual Assimilation Model (PAM) (Best, 1994, 1995). This model makes predictions about L2 discriminability based on the mapping between L2 phonetic contrasts and native L1 categories. Particularly, there are four possible mappings under this model. Ranging from mappings leading to the best discrimination to the worst, they are: Unassimilated (U), 2-category assimilations (2C), category-goodness assimilations (CG), and single-category assimilations (1C). Unassimilated contrasts are those that do not map to any native-language categories, such as Zulu clicks for English speakers (Best et al., 1988). Presumably since no native language categories are recruited the listener can rely on pure acoustic discrimination, which should be optimal. 2C assimilations are ones where each phoneme in the L2 maps to a unique phoneme in the L1, thereby allowing the user to discriminate the sounds using native phonology. Both CG and 1C assimilate the L2 contrasts to a single L1 category, but with CG the stimuli are perceived as varying in their goodness of fit, whereas for 1C they are perceived as equally good exemplars of the category. The varying degree of assimilated contrast discriminability has been supported by various studies including, prominently, Japanese-English approximant perception (Best & Strange, 1992), French-English approximant perception (P. A. Halle, Best, & Levitt, 1999), and English-Zulu fricative perception (Best, McRoberts, & Goodell, 2001), among others. It is important to note that all these predictions are based on some measure of the mapping between L2-L1 categories, yet no explicit metric of measuring

phonetic/phonological distance is ever provided for establishing the nature of the mapping. In the aforementioned studies, the contrasts are chosen because based on the intuition of the authors, they expect them to fit into a certain assimilation bin. Then, the intuition is confirmed with an identification experiment and the discrimination is examined to see if it conforms to the predictions made by the model. In the end, the correlation is only between behavioral patterns and lacks any formal motivation. There is no explicit model that can quantify the relationships and make accurate predictions about fine-grained discriminability of contrasts in a given L2 in relation to the native categories of a particular L1.

2.2.2 Learning, Eventual Attainment, and the Speech Learning Model

A classic theory of L1 language learning states that there is a critical period between birth and puberty during which we are able to acquire our native language, and language learning cannot be complete once the period is over (Penfield & Roberts, 1959). By most accounts, second language learning is no different, and learners can never attain native-like proficiency in a foreign language if they start to learn it after the critical period (Lenneberg, 1967; Johnson & Newport, 1989, inter alia) While there is disagreement as to the exact age and whether there are different critical periods for learning phonology, syntax, and morphological rules, there is widespread consensus that native proficiency will not be attained. Further, there is largely agreement that native language interferes with the second language. However, the exact nature of this interference and the nature of the learning appa-

ratus that the learner brings to the task of L2 learning is not a settled matter. Some researchers believe that adult learning of an L2 has to piggy-back on the phonetic and phonological categories and building blocks of the L1 and employ strategies to adapt where necessary (e.g., Best, 1994). Other researchers will admit the learned structure and rules of the L1 as a starting point but allow for the same learning apparatus as available for the L1 (e.g., J. E. Flege, 1995). These models, while differing in some assumptions, make testable predictions on discrimination abilities for phonetic contrasts in an L2 and largely agree on the initial state of the problem, where the discriminability of an L2 contrast can be determined from some mapping between the categories of the L2 and the L1. Some of these models go beyond initial perception and predict the eventual learnability as well as rate of learning of the contrasts as well (e.g., J. E. Flege, 1995).

A prominent model dealing with L2 perception and acquisition is the Speech Learning Model (SLM) (J. E. Flege, 1995). Unlike the PAM, the SLM makes predictions not only for initial naive perception but also for the ability to acquire, and the rate and quality of acquisition of non-native contrasts based on their mapping to existing L1 sounds. The SLM states that adult L2 learners are also capable of learning new contrasts, but this ability is qualified by the shared phonological space with L1 (hence bi-directional influence between L1 and L2) and depends on time and quality of input. A core testable prediction of the SLM is that the greater the dissimilarity of an L2 sound from the closest L1 counterpart, the greater the chance that a new category will be formed for the sound. When a new category is not formed, the L2 is assimilated with the closest L1 sound, leading to an intermediate representa-

tion over time. When a new category **is** formed, it will sometimes be dissimilated to make the contrast easier to perceive and produce, in a sense "overshooting the target". Similarly to the PAM, the SLM does not make any specific claims as to the nature of the perceptual distance between L1 and L2 sounds. There is also no quantifiable point at which one would expect to switch between assimilation and dissimilation. The SLM is explicit, however, in stating that this is a continuum and not an all-or-nothing phenomenon, so at least it allows for some transition between the binary states of assimilation or dissimilation. Also, while the SLM does make the claim that category formation will be allowed or blocked as well as the degree of accuracy of the eventual category, it is not specific in terms of the nature of the change over time and exactly how the category formation proceeds or at which point a separate category is present when a new category is in fact formed.

2.3 Seeking a Common Framework

My goal is to examine the nature of categorical effects in both L1 and L2 using a common model. In both contexts, the categories that the listener possesses inform their accuracy of perception. Operating in the same phonological space, L1 and L2 categories both affect perception of a sound, and there is no sense in which a sound is generated from an L1 or L2 outside the inference procedure. Discrimination then is a consistent process, with the part changing being which categories the listener has. For identification, we would of course be considering a different set of categories that can be used in the experiment since a naive listener of a foreign language only

has their native categories while a learner also has the categories of the second language. But, in either case, the categories that the listener possesses drive the perception behavior. It is an open question as to whether people represent L1 and L2 categories in the same way and whether those categories have the same effect on perception. In this work, I take the point of view of Flege and consider adult L2 learners to have access to the learning mechanisms of a child and hence to be able to eventually form similar categorical representations. This does not mean that they are quick learners nor that their representations are accurate or resemble native speakers. Rather, it indicates that at the computational level, the categories influence the optimal inference mechanism for intended speaker productions in a similar way, independent of exactly how the category representations are neurally implemented. I propose that the same mechanism that leads to categorical effects in native language perception is responsible for the misperceptions that occur in non-native contrasts. Further, as new non-native categories come online with L2 learning, this mechanism can predict how perception changes and given the right input will eventually allow for more accurate perception of sounds of the L2.

2.3.1 Unifying L1 Perception

In L1 perception, my goal is to unify behavioral perception findings for stop consonants, fricatives, nasals, and vowels under a single model. I draw a link between the categories of the language and the related perceptual warping. This mechanism does not rely on the sounds being from the first language or from being along a

specific continuum. All that is needed to predict their perception is to know the acoustics of the sound and the structure of the categories of the native language. Further, the mechanism does not have to depend on the specific cues that are used to identify a sound. Of course for the purpose of a computationally tractable problem, I restrict the cues that are involved in the perception of any given phoneme, but the approach in no way depends on the nature of a given cue to perception. Hence, as I will show in the next chapter, we can discuss categorical effects broadly and show how they arise for several different phonemes with different properties.

2.3.2 Extensions to L2 Perception

In the L2 background section, I outlined some shortcomings of the models used to examine perception in that domain. The core issue is that various distance metrics that define distances between phonemes and categories and assimilation and discrimination are not formalized and quantified. Also, predictions are only made at the course level between actual L2 contrasts and overall effects on perception along the phonetic space are not considered. Additionally, assimilation patterns and their discrimination counterparts are defined based on behavioral data and not as a consequence of a theory of categorical effects on perception. It is with these shortcomings in mind that I would like to investigate the use of a Bayesian model of categorical effects in speech perception applied to a cross-linguistic situation. With a concrete model we can quantify the perception of the non-native L2 contrasts, validating the qualitative predictions of the PAM and SLM, and getting a detailed

view of the particular perceptual measurements and categorical change over time. Also, once we can validate that a model can quantify such findings for both naive perception as well as eventual categorical formation, we will have a basis for testing particular effects of relevant parameters, such as category center, variance, and prior probability of category, all of which will be discussed in the next section on the model I use for my simulations. Perhaps even more importantly than just showing that we are able to extract meaningful category parameters and get a better understanding of patterns in L2 perception and learning, we can unify a possible approach to investigating categories across L1 and L2 research. Just as I parameterized categorical effects across phonemes in L1 independently of their underlying cues, this could serve as a foundation for investigating perception more broadly without strictly considering phonemes based on different features or cues separately.

Chapter 3: Bayesian Model of Perception and L1 Simulations

In this chapter, I consider how the model can be used to extract detailed category parameters from behavioral data in native language (L1) perception. I use these parameters to show how a unified model can account for varying degrees of categorical effects for a range of phonemes including vowels, stop consonants, fricatives, and nasals. Before actually considering the individual phonemes, I present the Bayesian model in detail so that we can see how it correlates with the category parameters I use to parameterize categoricity in phoneme perception. Below, I start by considering the level at which I unify these effects.

3.1 Levels of Computation

It is not enough to say that I want to account for two sets of findings using a single model or to show how they represent the same underlying phenomenon. It is important to distinguish at which level of representation and processing I want to unify the results. The possible approaches map unto the three levels of computation that were proposed by Marr (1982): computational theory, representation and algorithm, and physical implementation.

1. Computational Theory: This level concerns the goals of the computation.

Particularly, what is the computation that is being accomplished and why is the computation being accomplished in the first place.

2. Representation and Algorithm: This level concerns the representations that would allow the computation to be performed and the algorithm by which these input and output representations can be related to each other.

3. Physical Implementation: This level concerns the actual physical realization of the relevant computation.

For understanding speech perception, each level has a unique, but related, contribution. It would be impossible to paint a full picture of speech perception, or categorical effects that warp the perceptual space, without explaining the phenomenon at each level and showing how they are inter-related. For example, the neural implementation for perceiving time between two aural events and the frequency of a steady state signal are different, even if they can be mapped to each other Jeffress (1948)[1]. Hence, a unified account of categorical effects wold not be found here. More generally, speech perception cannot be explain just by consulting the neural implementation level. This would be akin to trying to "trying to understand flight by studying only feathers: it just cannot be done" (Marr, 1982). If we want to find how categorical effects for speech perception arise from a singular underlying cause we have to look to a higher level of analysis. We consult the computational level, which informs us as to the nature of the computation. At this

[1]While the original model focused on time of arrival differences between the ears to spatially represent sound source, the general method of mapping time differences to a spatial code can be used to directly relate continua such as VOT and formant space

abstract level, we can find a common cause for various categorical effects. Of course, looking at the computational level yields something akin to a first approximation of the full nature of categorical effects. At this level, we consider the essentials of the computation and ignore questions of intermediate productions, algorithms, and resource issues. To eventually get a full picture, one would have to consider the algorithmic and implementational levels as well. Specifically, without knowing how a strategy is implemented, we have no way of verifying if a proposed computation can even be carried out by the system that we are purporting to explain.

In the experimental and computational work that I present in this dissertation, I will be exploring the possibility of a cohesive underlying model purely at the computational level. The focus will be on finding the optimal solution to the problem of perceiving the speech sound produced by the speaker. My solution takes the form of computing the optimal inference of the intended speech sound, thereby addressing the target of the computation. There will not be a claim to any specific algorithm, set of representations, or neural implementations for this optimal strategy. Part of the beauty of such a model at the computational level is that it allows for the possibility of varying algorithmic and implementation levels of analyses for different sets of sounds. Specifically, for different sounds, the relevant continua that I am considering differ in terms of the underlying features, both acoustically and phonemically, that they represent. In fact, the neural implementation is almost definitely not the same for the perception of voice onset time in stop consonants and frequency bands of noise concentrations for vowels and fricatives. What is important here is not how we arrive at the representation of the continuum values, but rather what is the

computation that we perform over these values.

3.1.1 Why Computational Modeling?

There are numerous advantages to using computational models to investigate speech perception. One mentioned already is the ability to abstract away from different cues used for different phonemes and consider the problem as a general cognitive mechanism. Another also mentioned is the ability to abstract away from questions of neural implementation and consider speech perception as a case of optimal behavior. However, there are other very practical advantages with working with a formal computational model. Primarily, a formal model allows us to explore the effects of varying parameters in a certain framework and consider what would happen if they were adjusted one way or another. This allows me to test my hypothesis about speech perception by being able to simulate different real-life and hypothetical situations and seeing if the model behaves as expected. Running new "experiments" using a formal model is very low cost and has a quick turn-around, especially compared to studies with human participants, not to mention that results are clear and exactly reproducible. Also, the scale of experiments using formal models can be larger than that of human studies. Going beyond testing and experimenting with the model itself, once we have a certain amount of faith in the model we can use it to make novel predictions for how a human would behave in situations for which we have no behavioral data. This in turn can inform the most fruitful avenues for experimental research by identifying critical variables, particular

interactions, or likely underlying structures.

Some researchers do not consider the distinction between levels of analysis to be fruitful, whether we consider levels proposed by Marr (1982) or Newell and Simon (1976). Sun, Coward, and Zenzen (2005) believe that these separations are arbitrary and have failed to produce any major useful insight in cognition while adding unnecessary terminology to our theoretical apparatus. While I do not agree with their assessment, there certainly are broader advantages that go beyond the use of computational-level models and apply to formally specified computational models of cognition, where computational does not refer to a level of analysis but rather to the notion of using mathematically rigorous simulations to analyze and investigate human cognitive behavior. Behavioral tests with humans can only probe a superficial set of features, leaving us to either try to control for or ignore as random many possible sources of variances and individual differences. We can't just experiment our way to a full understanding of the language system or the human mind, we need some road guides, or sherpas. Formalizing our assumptions and giving us the ability to make a priori predictions about the nature and inner workings of human cognition to inform behavioral tests is critical (Sun et al., 2005). Such models can provide conceptual clarity, and provide easily testable, and refutable, predictions. According to Hintzman (1990), formal models also allow us to avoid common pitfalls in trying to reason backward from behavior to underlying processes, exploring and often refuting our assumptions about how combinations of processes would behave together. Formal models can also become theories in their own right, providing a way of stating complex theories of cognition and providing insight that a purely

theoretical or behavioral approach cannot (Sun, 2008). For a more in depth analysis of the role of formal computational models in cognitive sciences more broadly, see McClelland (2009) and Sun (2008).

3.2 The Bayesian Model

In order to investigate whether it is possible to model the effects seen so far as a qualitatively similar process, it is necessary to adapt a model so that a single framework can be used for all the different phonemes. For the purposes of the current paper, I will discuss previous work, recent adaptations, and new simulations in the context of a model originally developed by N. H. Feldman et al. (2009). This is a Bayesian model that was originally developed to explain the PME along the /i/-/e/ continuum. By expanding the model I am able to apply it to a broader range of data, including stop consonant and fricative perception. Here I present the details of the model and relevant theoretical and mathematical framework before proceeding to simulations.

The original model lays out a computational problem that the listener is solving as they perform identification and discrimination of sounds. It assumes that the listener is making a choice among a set of categories while listening to sounds coming from the continuum between these categories. The model attempts to formalize the assumptions that the listener is making about the generative process that produced the sounds as well as considering how these assumptions affect their perception of the sounds. The model formalizes both the identification and discrimination pro-

cesses discussed in the phoneme perception sections above, which I will consider once I have set up the relevant parameters. A graphical representation of the model appears in Figure 3.1. For presentation purposes and because of the nature of these particular studies, I restrict my attention to the case of two categories throughout the rest of this paper.

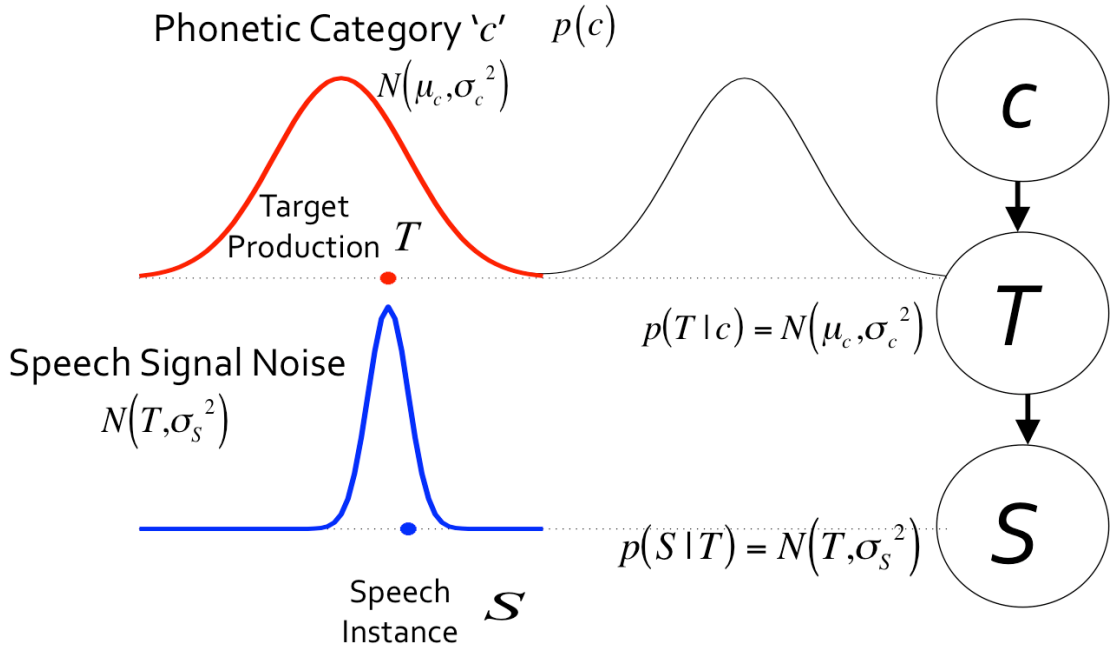


Figure 3.1: Full generative model for the production of sounds, representing the process the listener is assuming in terms of making their inferences of the underlying category and intended target production

I start with the listener's knowledge of the two categories, which I call c_1 and c_2. The next steps concern the process that the listener presumes is going on. First, the speaker chooses one of the two categories, which I will simply call c. I call this category our 'underlying category'. This is not to be confused with the term 'underlying category' used in phonology to represent the phonemic category that underlies allophonic variation. Instead, I use 'underlying category' to refer to

phonetic category that has not been corrupted by noise in the speech signal. In our model this category is represented as a Gaussian distribution around the category mean μ_c with variance σ_c^2. There is nothing intrinsic to the model that requires this distribution to be Gaussian. This is largely chosen for conformity with previous work and mathematical elegance. It is important that the distribution be unimodal so that there is a coherent idea of warping toward the category center. For things like VOT, it is actually possible that a Gauss-exp distribution would work better since the distribution can only go so low on the VOT scale, but can stretch out quite a lot in the other direction, similarly to how reaction times are often modeled in psycholinguistic studies. However, for the current work, I restrict the model to Gaussian distributions.

I am agnostic to the particular dimension for the mean and variance, but in principle it can represent any cue used in phoneme identification. In practice, in my simulations, I consider dimensions including VOT (ms) for stop consonants, F1/F2 (Hz) for vowels, central frication frequency (Barks) for fricatives, and formant onset transitions for POA (Hz) for nasals. This mean and variance of the category being used in the generative procedure by the speaker is known by the listener from previous exposure to sounds from this category in the language. For the purposes of the model, I do not concern myself with how exactly such categories were learned, but rather the perception that occurs once the categories are already acquired. Hence, the mean represents the center of the category in perceptual space. The variance here is assumed by the listener to be derived from processes that provide useful information about the nature of the sound. This information

can include cues to speaker identify or dialect detection. It may also convey mood, intonation, and emotion. This variance may also reflect coarticulatory effects which allow the listener to predict upcoming sounds (Gow, 2001). Because of this, I call the categorical variance of the underlying category 'meaningful'.

The next step in the generative procedure is the selection of an intended target production from the normal distribution $N(\mu_c, \sigma_c^2)$. I call the intended target production T. In terms of probabilities of choosing a specific target production, T, from the distribution, $p(T|c) = N(\mu_c, \sigma_c^2)$. Once the intended target production is chosen, it needs to be articulated for the listener to hear a sound. This process introduces additional articulatory, perceptual, and acoustic noise that distorts the signal. This is formalized as an additional Gaussian distribution around the intended target production with mean T and variance σ_S^2. As before, the choice of a Gaussian distribution is largely for consistency and mathematical elegance, not for any deep theoretical consideration. I call this additional variance 'noise' variance since it doesn't provide useful information to the speaker and distorts the speech signal. We call the actual speech instance that the listener perceives S and it is sampled from this distribution, with probability $p(S|T) = N(T, \sigma_S^2)$. We can also consider the overall distribution of possible speech sounds related to the underlying category chosen by the speaker at the beginning of the generative procedure. If we integrate over all possible intended target productions, T, then we can describe the distribution of speech sounds as $S|c = N(\mu_c, \sigma_c^2 + \sigma_S^2)$.

Now I can consider how this model relates to the behavioral tasks described in the phoneme perception sections above. In those tasks, listeners are asked to either

identify the category of the sound or to tell if two sounds are the same or different. In the model, the identification task corresponds to retrieving the underlying category, c. The discrimination task maps to recovering the intended target production for each of the stimuli heard by the listener and then comparing them to see if they are the same or different. The listener is presumed to be recovering both the phonetic details about the target production as well as category choice information when they perceive sounds. By fitting the model to behavioral tasks performed by listeners, I find the optimal setting of parameters that best fits the data. I then examine how these parameters relate to the degrees of categoricity seen in the perception of different phoneme continua.

Let us consider how the listener retrieves this information that they need. First, note that the listener does not have access to the intended target production, T, that the speaker meant to articulate. However, as a speaker of the language, the listener does have their own representation for the underlying categories, $N(\mu_{c1}, \sigma^2_{c1})$ and $N(\mu_{c2}, \sigma^2_{c2})$, and the noise variances associated with that particular phonetic category, σ^2_S. They also have access to the actual speech sound that they perceived, S. This means that the listener will use a combination of actual perceived speech information and knowledge of underlying categories in inferring what they heard. This relationship between the contribution of S and μ_{c1} and μ_{c2} will become important as we move forward in evaluating the varying effects of categories in identification and discrimination tasks. In the model, the identification task corresponds to finding the probability of a particular category given the perceived value along the dimension of interest. In other words, it entails computing $p(c|S)$. The discrimination task

corresponds to inferring the target production for each of two perceived stimuli and deciding if they are the same or different. This required finding the probability of a given target production given the perceived value along the dimension of interest, or computing $p(T|S)$. For both of these inference procedures I turn to Bayes' Rule, which I discuss below along with the relevant computations and how they relate to the modeling throughout the dissertation.

3.2.1 Introducing Bayes' Rule

Bayes' Rule is derived from a simple identity in probability theory (Bayes, 1763). In terms of calculating a belief in a hypothesis based on some observed data, Bayes' Rule states that the posterior probability of a hypothesis given some data can be calculated from the probability of the data given the hypothesis multiplied by the prior belief in the hypothesis and then normalized by the overall (marginal) probability of the data given all possible hypotheses (Equation 3.1).

$$p(h|d) = \frac{p(d|h)p(h)}{\sum_h p(d|h)p(h)} \tag{3.1}$$

The denominator on the right-hand side of the equation is just the marginalized representation of the overall probability $p(d)$. The term of the left, $p(h|d)$ is called the posterior probability. The term $p(d|h)$ is called the likelihood, since it describes how likely the data is under a certain hypothesis. The final term, $p(h)$ is called the prior, since it is the probability of the hypothesis (ie: our belief in the hypothesis) before we saw the data. Often these prior probabilities are uninformed and set using

some heuristic, or they are just uniformly distributed among all possible hypotheses, as is the case in some of our simulations below.

The setting of the hypothesis and data will be different depending on what behavior I am modeling (identification vs discrimination). In the section below I go through in detail how the inference procedure for the listener is structured, and what parameters I can extract by fitting the model to the listener's behavioral data.

3.2.2 Bayes' Rule for Identification

First I consider the behavioral task of identification. For a listener, the task of identifying a sound involves picking the correct category label for the sound. In the generative model, this means inferring c from the speech sound S. To do this I set the relevant variables in Bayes' Rule as follows:

- Data: speech sound, S
- Hypothesis: the underlying category, c
- Likelihood: the probability of hearing the speech sound given the category, $p(S|c)$
- Prior: the probability of choosing the category, 0.5 (uniformly distributed between two possible categories)

Given the above, I calculate the posterior probability $p(c|S)$ which the listener needs to infer to perform an identification task during behavioral trials. Bayes' rule for identification is:

$$p(c|S) = \frac{p(S|c)p(c)}{\sum_c p(S|c)p(c)} \tag{3.2}$$

If I rewrite using the actual probability distributions I will be able to see what the critical parameters are here that I can recover via model fitting. Afterward I will go on to do a full derivation of the actual values that I will use for fitting. The expression for what I am calculating in identification, using probability distributions from the model, is then:

$$p(c_1|S) = \frac{N(\mu_{c1}, \sigma_{c1}^2 + \sigma_S^2)(0.5)}{(0.5)N(\mu_{c1}, \sigma_{c1}^2 + \sigma_S^2) + (0.5)N(\mu_{c2}, \sigma_{c2}^2 + \sigma_S^2)} \tag{3.3}$$

I make the simplifying assumption that both categories are equally likely before any speech sound is heard, substituting 0.5 for p(c) in this equation. In other words, I am not taking into account different phoneme frequencies. This is not a realistic assumption for speech processing since people have shown that vowel frequencies (Gimson, 1980; Wioland, 1972; Fok, 1979; B. Fry, 1947) and consonant frequencies (Crawford & Wang, 1960; Mines, Hanson, & Shoup, 1978) differ greatly. While this is an issue for lexical identification, where phoneme frequencies play a big role, it is possible that these effects are diminished in a laboratory setting where participants merely choose between two particular phonemes. While this is something that can be investigated, I nonetheless proceed with this simplifying assumption.

Notice that the values that appear in this equation are: (1) μ_{c1}, (2) μ_{c2}, (3) $\sigma_{c1}^2 + \sigma_S^2$, and (4) $\sigma_{c2}^2 + \sigma_S^2$. It is these values that I would be able to recover via a fit of our model to the behavioral data produced by the listener. In the simulation

section I will see how this fits into the overall process of extracting parameters via model fitting, and how these parameters guide our understanding of the gradient effects of categoricity.

Now I need to derive the actual equation that I will fit to the data. In the original application of this model to vowel data in N. H. Feldman et al. (2009), there was a simplifying assumption that underlying category variances for the two categories $c1$ and $c2$ were equal. This meant that only one sum of variances would need to be considered, $\sigma_c^2 + \sigma_S^2$. However, this turns out to be a very inaccurate assumption for stop consonants. It has been shown that voiced and voiceless stop consonants have substantial differences in their variances of voice onset time (VOT) (Lisker & Abramson, 1964a). As a result, the present model is extended to allow for different variances for the two categories. The final formula for the probability of the sound having been generated from category 1 given the actual perceived stimulus is presented in Equation 3.4. Full derivations for both versions of the model are provided in Appendix A.

$$p(c_1|S) = \frac{1}{1 + \sqrt{\frac{\sigma_1^2}{\sigma_2^2}} \times exp\frac{(\sigma_2^2-\sigma_1^2)S^2+2(\mu_{c_2}\sigma_1^2-\mu_{c_1}\sigma_2^2)S+(\mu_{c_1}^2\sigma_2^2-\mu_{c_2}^2\sigma_1^2)}{2\sigma_1^2\sigma_2^2}} \tag{3.4}$$

where $\sigma_1^2 = \sigma_{c_1}^2 + \sigma_S^2$ and $\sigma_2^2 = \sigma_{c_2}^2 + \sigma_S^2$.

I can now find the optimal fit of this model to the underlying behavioral data. I do this by computing an error function between this model and the behavioral data and then running an error minimization routine in MATLAB to find the optimal parameters. The process of deriving the error function, finding initial parameter

settings, and running the error minimization routine will be discussed in Section 3.4.

3.2.3 Bayes Rule for Discrimination

Now I can consider the behavioral task of discrimination. For the listener, the discrimination task involves inferring the most likely value of the target production, T, for a pair of stimuli and then comparing the values to see if the intended target productions were the same or not. The further apart the pair of Ts are judged to be, the higher the chance that the person will decide that the stimuli are different. The relevant variables for the inference, in terms of Bayes' rule, are as follows:

- Data: speech sound, S
- Hypothesis: the target production, T
- Likelihood: the probability of hearing the speech sound given the target production, $p(S|T)$
- Prior: the phonetic category structure, $p(T|c)$

Given these values I can calculate the posterior probability of $p(T|S)$ which the listener needs to infer to perform the discrimination task during behavioral trials. First, since the target production, T, could have derived from either underlying category, I do a weighted sum over the two categories, where the weighting term is the probability of the category being the underlying one chosen given the speech

sound heard, $p(c|S)$, which I computed above in the identification section. The posterior has the form:

$$p(T|S) = \sum_{c} p(T|S,c)p(c|S) \tag{3.5}$$

Now, I can compute the posterior for a specific category, $p(T|S,c)$, using the values introduced above. Bayes rule for discrimination looks as follows. Note that I have replaced the summation term with an integral term since unlike the category variable in the identification task, the target production is a continuous variable.

$$p(T|S,c) = \frac{p(S|T)p(T|c)}{\int_T p(S|T)p(T|c)} \tag{3.6}$$

If I rewrite using the actual probability distributions we will be able to see what the critical parameters are here that we can recover via model fitting. Afterward I will go on to do a full derivation of the actual values that I will use for fitting. The expression for what I am calculating in discrimination, using probability distributions from the model, is then as follows in Equation 3.7.

$$p(T|S,c) = \frac{N(T,\sigma_S^2)N(\mu_c,\sigma_c^2)}{\int_T N(T,\sigma_S^2)N(\mu_c,\sigma_c^2)} \tag{3.7}$$

Plugging this back into equation 3.5 and expanding the summation term we get:

$$p(T|S) = \frac{p(c_1|S)N(T, \sigma_S^2)N(\mu_{c1}, \sigma_{c1}^2)}{\int_T N(T, \sigma_S^2)N(\mu_{c1}, \sigma_{c1}^2)} + \frac{p(c_2|S)N(T, \sigma_S^2)N(\mu_{c2}, \sigma_{c2}^2)}{\int_T N(T, \sigma_S^2)N(\mu_{c2}, \sigma_{c2}^2)} \tag{3.8}$$

Notice that the values that appear in this equation are: (1) $p(c_1|S)$, (2) $p(c_2|S)$, (3) T, (4) σ_S^2, (5) σ_{c1}^2, and (6) σ_{c2}^2. Out of these values, the noise variance, σ_S^2, is the only one that needs to be fit by the model. This is because the first two, $p(c_1|S)$ and $p(c_2|S)$, are known from the identification part of the model fitting, T is the value being calculated, and the last two, σ_{c1}^2 and σ_{c2}^2, can be calculated by subtracting σ_S^2 from the two sums of variance terms inferred in the identification stage above. We also need a ratio constant, which we will call K, to relate the model discrimination predictions to the discriminability metrics used in the behavioral experiments, particularly d' scores which are used as more valid measures than pure accuracies. This additional parameter is merely a constant term that stretches the range of the model value to the range of behavioral findings, and does not provide critical information about the structure of the problem. Hence, we have 2 free parameters in the model that will be estimated via fitting the behavioral data: σ_S^2 and K.

Now I need to derive the actual model equation that I can fit to the discrimination data. This means calculating the optimal value for the posterior distribution for the target production given the speech sound. In this case, the calculation corresponds to the mean value of the posterior distribution, which is a mixture of Gaussians obtained from the Gaussian prior, $p(T|c)$, and the Gaussian likelihood,

$p(S|T)$, via Bayes rule. The base equation can be seen in Equation 3.9, with the expanded form for the case of different category variances in Equation 3.10. The full derivation can be found in Appendix B.

$$E[T|S] = \sum_c p(c|S) \frac{\sigma_c^2 S + \sigma_S^2 \mu_c}{\sigma_c^2 + \sigma_S^2} \tag{3.9}$$

$$E[T|S] = p(c_1|S) \frac{\sigma_{c1}^2 S + \sigma_S^2 \mu_{c1}}{\sigma_{c1}^2 + \sigma_S^2} + p(c_2|S) \frac{\sigma_{c2}^2 S + \sigma_S^2 \mu_{c2}}{\sigma_{c2}^2 + \sigma_S^2} \tag{3.10}$$

In the simulation section I will see how this fits into the overall process of extracting parameters via model fitting, and how these parameters guide our understanding of the gradient effects of categoricity, but for now I want to point out the relationship between the meaningful variance, σ_{c1}^2 and σ_{c2}^2, and the noise variance, σ_S^2. After fitting the discrimination curves, we have the independent contribution of these two parameters. It is the ratio of the meaningful to noise variance that I will show to correspond with the degree of categorical effects in phoneme perception.

The expected value for the target production is a weighted average of contributions from each possible underlying category. The contribution from each category is itself a linear combination of the speech sounds, S, and the underlying category mean, μ_c. Note here that the acoustic value of S is being multiplied by the meaningful category variance term, σ_c^2. A higher meaningful variance term means a greater contribution from S in the inferred target production. Meanwhile, the acoustic value for the category mean, μ_c, is multiplied by the noise variance term, σ_S^2. This means that a higher noise variance term leads to a greater contribution from μ_c,

the category mean, in the inferred target production. As a result, varying the ratio between these two variance terms will lead to varying influences from the category mean and the speech sound, effectively controlling categorical effects. Motivation for, and a further explication of, this ratio is presented in section 3.3.

What is left is to find the optimal fit of this model to the underlying behavioral discrimination data. As with the identification fitting, this was done by computing an error function between this model and the behavioral discrimination data and then running an error minimization routine in MATLAB to find the optimal parameters. The process of deriving the error function, finding initial parameter settings, and running the error minimization routine will be discussed in Section 3.4.

3.3 The Tau Ratio

I stated at the end of the previous section that the ratio of the meaningful category variance to the noise variance parameter is related to the proposed continuum of categoricity effects for the different phonemes that I am considering. First I consider how it is that these two values contribute to varying degrees of effects of categories. I stated that listeners are inferring the target productions of the speaker through a signal that is inherently noisy. In the process of doing so, the listener has access to the speech sound, S, which represents the actual acoustic cues to what the sound is, as well as the underlying category mean μ_c, which represents their knowledge of categories which they can use to guide their inference since they know that speakers tend to produce sounds near category centers. The amount of mean-

ingful and noise variance guides them in terms of which knowledge they can rely on more in making the inference. If more of the variance is meaningful, then that means they should rely more on acoustic details in perceiving the sound, hence giving more weight to S. However, if much of the variance comes from noise, then that renders the acoustic details unreliable, and the listener instead uses their pre-existing knowledge of categories to guide the inference, giving more weight to μ_c. It is exactly this relationship that I examine by considering the ratio of meaningful:noise variance, which I will call τ, or τ_c for a specific category c.

We can gain insight into the continuum of τ values by looking at the extremes. As τ approaches zero, either the meaningful variance of a category approaches zero, or the noise variance grows infinitely large, leading the user to depend entirely on their existing knowledge of categories. Under these conditions, perception would look extremely categorical, as the user throws out any contribution of fine acoustic detail and instead only uses the probability of belonging to a category based on the category mean. At the other extreme, as the ratio approaches infinity, either the meaningful category variance grows large or the noise variance goes to zero. In both cases, the contribution of the underlying category means shrinks to nothing and perception is guided purely by the details in the speech stimulus, S. This means that perception would be entirely continuous and veridical to the acoustic signal. Overall, this relationship represents the degree to which perception is biased by the effect of category membership, and can account for the gradient effects of categoricity we observe in various behavioral tasks.

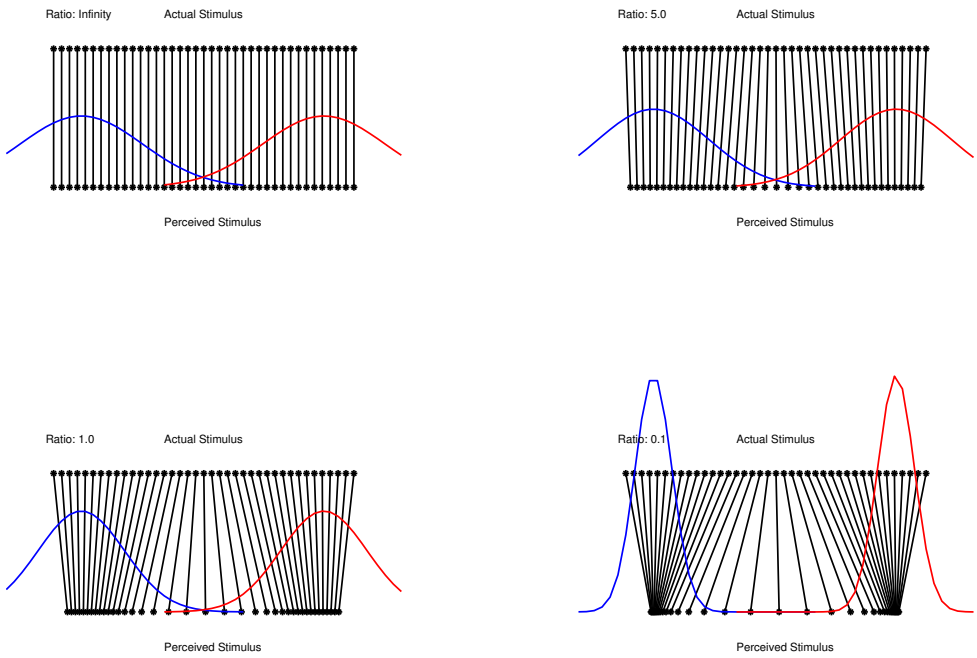


Figure 3.2: These simulations illustrate the effect of varying the ratio of meaningful to noise variance. Warping from actual to perceived stimuli is shown in the dispersion of the vertical bars toward category centers. Total variance is held constant throughout the simulations, with the amount of variance attributed to underlying category variance shown in the two Gaussian distributions overlaid over the perceptual warping bars. Ratios presented include: (a) infinity, (b) 5.0, (c) 1.0, and (d) 0.1

3.3.1 Degrees of Warping

The degree of warping along a given continuum can be visualized graphically by connecting the actual acoustic values of the stimuli on top with the perceived values on the bottom. This process provides an appealing visual perspective of the degree of warping for any given value of τ. In Figure 3.2 we see what happens as we move from a situation with no noise (ratio of infinity) to a condition where

there is ten times more noise than meaningful variance (ratio of 0.1). Along with the warping of actual to perceived stimuli, each chart is overlaid on top of the categorical variances that make up the meaningful variance component of the overall variance (I hold the sum of the meaningful and noise variances constant in these simulations).

In the top left graph, with no noise, all stimuli are perceived veridically and we have no warping of the signal. This is the extreme case with an infinite ratio, but instructive since this is what would happen if we had full faith in the acoustic signal. In terms of our model, this means that the contribution of the perceived speech stimulus, S, completely trumps the mean of the underlying category, μ_c. This graph also serves as a good reference point, with the gaussian distribution in this case the same as the sum of the two variances in all other graphs (i.e. since there is no noise, the entire variance is due to the meaningful variance). We then see what happens in cases where more and more of the variance is attributed to perceptual noise, thereby shrinking the meaningful:noise variance ratio. Since we hold the sum of the variances constant, greater noise variance means a much smaller meaningful variance, and consequently we see the gaussians in the graphs get tighter around the category centers as we reduce the ratio. More importantly, with the falling ratio we clearly see greater and great warping as the individual stimuli get pulled more and more strongly toward the centers of the categories. At the last step, with a ratio of 0.1, only the very central stimulus is perceived veridically, with immediate strong pulling toward category centers on either side, and many of the stimuli almost completely sucked in by the category center. Moving down on the τ continuum represents the relative size of the meaningful category variance getting

smaller and smaller. In terms of the model, this means that the contribution of the perceived speech stimulus, S, is decreased compared to the contribution of the mean of the underlying category, μ_c. As this happens, the perceived stimuli move closer and closer to the center(s) of the underlying category(s), μ_c. In terms of motivation of why this might happen, we can consider the explanation above in terms of perception in a noisy channel. The meaningful variance going down means that there is a greater contribution from the noise variance. The listener can't rely on the information coming in through the noisy channel. Instead, they must rely on their prior knowledge of categories. As a result, the perceived stimuli get pulled into the centers of the categories, of whose structure the listener has prior knowledge.

These pictures, along with the motivation presented, show that the ratio of variances can represent the degree of categorical effects on perception. What is left to show is whether the phonemes that I have been considering thus far correspond to this continuum. I will need to fit the model to each type of phoneme's identification data and discrimination data to get the relevant ratio of meaningful:noise variance. Then, I can compare these ratios to the warping continuum that I just described. If the phonemes map onto the τ continuum in a way that correlates with the respective behavioral findings, then the model captures the relevant behavioral patterns via a single parametric variation, and there is no need to appeal to two independent effects to describe categorical effects for vowels and consonants.

3.4 Fitting to Data

In this section I present simulations for vowels, stop consonants, fricatives, and nasal consonants using the same model. I show that the model is able to provide a good fit to the behavioral data with an appropriate setting of parameters, and further, that the derived parameters are precisely of the type that yield the proposed single qualitative source of categorical effects proposed above. Throughout this section I focus on two aspects of the model fit: the actual fit of the model to the data (in the sense that it accounts for the identification and discrimination curves in an adequate manner) and the derived τ value (particularly, where it falls on the continuum relative to the other phonemes). For all of my simulations, whether previous or new work, I will be following the procedure described in Table 3.1, with details in the next section.

3.4.1 Model Fitting Steps

3.4.1.1 Setting a Category Mean

First I set the mean for one of the two categories. This is because the model is otherwise underspecified; an infinite number of category means provide for the same fit to the identification data (i.e. the parameters are not identifiable). Technically, this is because fitting the model derives the equivalent of the sum of the category means, so moving one mean up would just move the other one down. Since a specific set of means are needed in order to get the proper variance ratio in the following

Simulation Steps for Model Fitting		
Step	**Description**	**Derived Parameters**
1	Set μ_{c1} on the basis of production data	μ_{c1}
2	Determine μ_{c2}, σ_1^2, and σ_2^2 from identification data using Equation 3.4	μ_{c1}, μ_{c2}, σ_1^2, σ_2^2, giving us the full category structure of the perceived categories
3	Determine the ratio of the meaningful category variances, σ_{c1}^2 and σ_{c2}^2, to noise variance, σ_S^2, by fitting acoustic differences between percepts, $E[T\|S]$, in the model (Equation 3.9) to a distance measure such as d'.	μ_{c1}, μ_{c2}, σ_{c1}^2, σ_{c2}^2, σ_S^2, giving us the category structure of the underlying categories
4	Compute τ from the meaningful category and noise variances and examine where they fall on the continuum presented above.	τ, giving us the degree of warping along our derived continuum

Table 3.1: Steps used to complete simulations of the model to behavioral data for vowels, stop consonants, fricatives, and nasals[2]. (NOTE: $\sigma_1^2 = \sigma_{c1}^2 + \sigma_S^2$ and $\sigma_2^2 = \sigma_{c2}^2 + \sigma_S^2$)

steps, we fix one of the means. In my simulations I set the mean based on production data from native speakers.

3.4.1.2 Identification Fitting

To fit the model's identification predictions to the behavioral data, I use Equation 3.4. With the setting of one of the category means in the previous step, I am left with three parameters to set via the fitting. What I can now do is derive an error function between this equation and the value of the actual values found through the behavioral study. I use mean squared error (MSE) to find the set of parameters σ_1^2, σ_2^2, μ_{c1}, and μ_{c2} that give the optimal fit. Once the identification is

[2]The procedure for fitting nasals is actually different in step 3 since instead of d' data we have to fit to raw accuracies. This step will be replaced by a Monte Carlo simulation and detailed in Section 3.4.5.1

fit, I have in effect inferred the sum of the variances that the listener in our generative model assumes, leading to the structure of the Gaussian distributions for $p(S|c) = N(\mu_c, \sigma_c^2 + \sigma_S^2)$. This is the structure of the perceived category, as opposed to the underlying category.

3.4.1.3 Discrimination Fitting

Now we need to pull apart the contribution of the meaningful category variance and the articulatory and perceptual noise variance to the overall variance derived in the previous step. Note that the discrimination equation also has the separate parameter σ_S^2. This, in effect, is the only free parameter that is being inferred in this step. Separating out the noise variance serves a dual purpose. First, it allows us to calculate the ratio that I show to correspond with degree of categorical effects. Second, it provides an independent test of the models' ability to fit behavioral data. While in the identification fitting section I set one of the means in order to constrain the underspecified model, here I do not set any further parameters. This allows us to confirm that the model fully captures the patters observed in human behavior.

Once this step is complete we can see the shape of the underlying category distributions that are known to the listener in the generative model. We then check if the category structures found by the model correspond well to production data. This is especially relevant for cases where the distributions of the phonemes at the two ends of the continuum are non-parallel to each other, exhibiting very different variances and possibly possessing very different perceptual signatures in

terms of participants' ability to discriminate stimuli within the category. In the sections below I will draw on comparisons between model and production data where available and appropriate.

3.4.1.4 Calculate the Variance Ratio τ

As the final step of the simulation process, I compute the value of τ for the phoneme continuum under consideration. τ specifically quantifies the strength of the pull toward category centers. Mathematically speaking, this calculation is trivial given the previous three steps. We just divide the meaningful category variance, σ_c^2, by the noise variance, σ_S^2. Given, however, that I am fitting the categorical variance separately for each phoneme, technically we get two values of τ; τ_{c1} and τ_{c2}. Therefore, its not technically correct to say that our values of τ characterize warping along the continuum between the two phonemes. Rather, they characterize warping for either of the phonemes at the ends of the continuum. In effect, then, they characterize the warping for an idealized continuum if it were to consist of two identical categories at its ends. For our purposes, looking at the individual warping parameters for the phonemes is what we want since we want to be able to capture varying category structure and varying within-category discrimination. In my simulations, the τ values end up being very close for the phonemes at the two ends of the continuum for all phonetic categories except stop consonants, for which the within category variance for the voiced stop consonant /b/ is much lower than that for /p/.

Stimulus Number	F1 (Hz)	F2 (Hz)
1	197	2489
2	215	2438
3	233	2388
4	251	2339
5	270	2290
6	289	2242
7	308	2195
8	327	2148
9	347	2102
10	367	2057
11	387	2012
12	408	1968
13	429	1925

Table 3.2: Formant Values for Stimuli Used in the Multidimensional Scaling Experiment, as reported in Iverson and Kuhl (2000)

3.4.2 Original Vowel Fitting

N. H. Feldman et al. (2009) were the first to fit a version of this model to data from behavioral studies. In their model they made a simplifying assumption that stipulated that the two categories have equal variance. Their overarching goal was to consider categorical effects throughout cognition, and they took the PME as a demonstrative example of such effects. As a first step, they tested their account on a vowel continuum considered in previous research. They used identification and discrimination data from previous studies by Lotto et al. (1998) and Iverson and Kuhl (1995) respectively. In the end, after doing their own additional behavioral study to account for issues with the multi-dimensional scaling approach of the previous studies, they found a ratio of variances along the τ continuum that I use as a basis for comparison in further simulations.

N. H. Feldman et al. (2009) modeled data from a paper by Iverson and Kuhl (1995) that used a multi-dimensional scaling technique to examine discrimination performance by participants on the /i/-/e/ continuum. The formant values for the continuum are reproduced in Table 3.2 below, as reported by Iverson and Kuhl (2000). Even though the parameters here are given in Hertz, the continuum was based on equal sized steps in Mels, a psycho-acoustically accurate measurement scale for frequencies. The Mel scale was designed such that equal steps on the scale are based on difference limens (S. S. Stevens, Volkmann, & Newman, 1937). The discrimination data that was fit by the model was the output of the multi-dimensional scaling solution. Even though d' data was available, the MDS data represented the entire range of stimuli and therefore provided a broader range upon which the model could be fitted. For identification, however, the data from Iverson and Kuhl (1995) was insufficient, or at the very least not optimal. As was pointed out in Lotto et al. (1998), there was a discrepancy in how the identification and discrimination data were collected by Iverson and Kuhl (1995). The stimuli in discrimination trials were always presented in pairs. However, stimuli in the identification trials were presented in isolation. Because of known effects of context on perception, this meant that the category of the same stimuli might be being perceived differently in the two experiments. To circumvent this issue, identification data were taken from Lotto et al. (1998), who repeated the identification and discrimination experiments from Iverson and Kuhl (1995), but presented the stimuli for both experiments in pairs.

N. H. Feldman et al. (2009) showed the the model provides a close fit to the behavioral findings from Iverson and Kuhl (1995). However, it should be noted

that this was primarily a proof of concept that showed that the generative model is capable of fitting behavioral data. The actual derived parameters were judged to be inaccurate due to skewing introduced via the multi-dimensional scaling (MDS) procedure. To remedy this issue, Feldman et al. conducted a behavioral study that led to more appropriate values for the variances, and in turn for the τ ratio discussed previously. That study, as well as the derived parameters, are presented below.

N. H. Feldman et al. (2009) wanted to see how noise affects perception and whether the model captures this via the noise variance parameter as would be predicted. They conducted a behavioral study where they measured response times during same/different judgments for pairs of stimuli. The stimuli were the same ones used in Iverson and Kuhl (1995) and presented here in Table 3.2. Their experiment had a no-noise and a noise condition. From the data, they generated confusion matrixes for the stimuli in the experiment. Then, they wanted to see what setting of parameters in the model would best fit this confusion data. However, for their purposes, they modified the model slightly in order to have it predict same/different judgments directly. In effect, they were computing the probability that the distance between the inferred target productions was less than or equal to ϵ given the two perceived speech stimuli, S_1 and S_2. Since the same contrast was presented a total of n times during the experiment, the overall confusion of the two stimuli is measured by the binomial distribution $B(n, p)$, where p is the probability mentioned above and presented below in Equation 3.11.

$$p(|T_1 - T_2| \leq \epsilon \mid S_1, S_2) \tag{3.11}$$

The role of this parameter in the extended model is similar to that of the observer response criterion in signal detection theory (Green & Swets, 1966), where ϵ determined the size of the judged distance between stimuli necessary to yield a positive response, in their case a response of "Different".

In order to minimize free parameters in the model, they held constant all other parameters. To do this, they used the category means, $\mu_{/i/}$ and $\mu_{/e/}$, as well as the categorical variance, σ_c^2, from the MDS simulations. They found that the model was able to find very good fits to both conditions and, more importantly, that the noise parameter was independently a good predictor above and beyond the setting for the threshold parameter, ϵ. For my current experiment I am not interested in perception in the presence of additional noise, so I use the noise variance term from the no-noise condition, which was found to be $\sigma_S^2 = 878(\sigma_S = 30mels)$, as the appropriate noise variance to capture discrimination performance for the /i/-/e/ continuum.

The full set of parameters inferred through fitting the above basic and extended model can be found in Table 3.5 in the row for vowels. With the values for the meaningful and noise variance set to $\sigma_c^2 = 5,873$ and $\sigma_S^2 = 878$, the critical ratio for vowels is $\tau_V = \frac{5,873}{878} \approx 6.69$. The corresponding graphical warping picture for this τ value is shown in figure 3.6a, together with analogous graphs for other phoneme categories. We can see that there is not much warping between the actual and perceived stimuli. There is, however, some effect of categories, with stimuli closer to the categorical centers pulled together and slightly greater distances at the category boundary as the stimuli are pulled apart. This value and warping picture serve as baselines to consider behavioral data for other phonemes.

Stimulus (S_i) #	VOT (ms)	% Ident. of S_i as /p/	Hit	False Alarm	Computed d'
S_0	-50	0.0			
S_1	-40	0.0	0.33	0.25	0.24
S_2	-30	0.0	0.30	0.25	0.17
S_3	-20	0.016	0.30	0.30	0.01
S_4	-10	0.012	0.39	0.31	0.20
S_5	0	0.012	0.54	0.36	0.46
S_6	10	0.027	0.73	0.37	0.93
S_7	20	0.296	0.95	0.37	1.98
S_8	30	0.946	0.94	0.34	1.94
S_9	40	1.0	0.64	0.30	0.87
S_{10}	50	1.0	0.47	0.25	0.60
S_{11}	60	1.0	0.42	0.22	0.57
S_{12}	70	1.0			

Table 3.3: VOT for stimuli used in behavioral experiment, along with identification and discrimination data, for the /b/-/p/ continuum

3.4.3 Stop Consonants and Fricatives

For the simulations with stop consonants and fricatives the relevant measure was not an ϵ-distance same/different judgment but a measure of the psychoacoustic distance between stimuli. The relevant behavioral measure for this is d-prime (d'). In the model simulations, we treat d' as representative of perceptual distance between stimuli and hence proportional to the distance between two target productions. Because they are proportional and not identical, we fit for a scaling parameter, which is not informative for the representation of the categories and therefore not discussed. By fitting the d' metric we also do not have to have access to the original underlying data from the participants, which facilitates the fitting procedure for stop consonants, since it is based on data from a different lab.

3.4.3.1 Simulation 1: Stop Consonants

Now I turn my attention to stop consonants. As already reviewed, stop consonants have been found to exhibit very strong categorical effects in perception during behavioral experiments. In the model, this would possibly stem from a low meaningful category to noise variance ratio. As such, in the presence of noise, listeners would rely much more on their knowledge of underlying categories rather than the detail available in the acoustic stream. However, even if the relative contribution of category means and acoustic details does contribute to consonant perception, such factors as innate phonetic boundaries (Eimas et al., 1971) and auditory discontinuities (Pisoni, 1977) may continue to play a role. This additional effect on stop consonant perception would make it impossible to fully explain the behavioral findings in the terms that the model proposes. Because of this, a good fit to the behavioral data would be a particularly strong argument in favor of positing a unified account for categorical effects in phoneme perception.

For stop consonants I consider both identification and discrimination data derived from Wood (1976). Their experiments investigated the perception of stimuli between /b/ and /p/, varying along the voice onset time (VOT) continuum. Their stimuli were synthetically created along the continuum ranging from -50 to +70 ms VOT. Their identification task was a classic forced identification task. For discrimination, they administered both a 10-ms and 20-ms difference AX discrimination task, in which participants heard one stimulus, A, and then had to decide whether the second stimulus, X, was the same or different as the first. For the simulations

below, I used data from their 20-ms discrimination condition since the 10-ms discrimination scores were prone to floor effects, being so close together that discrimination was almost impossible, with only one value (across the boundary) reading a d' score above 0.5. Values for both identification and discrimination can be found in Table 3.3. I use d' as the measure of perceptual distance, which I computed from "hit" and "false alarm" values reported in a graphic representation in their paper [3]. No correction was applied to the d' calculation since none of the false alarm or hit rates were close to zero or one in the original findings in Wood (1976).

As a first step, I need to set the mean for one of the categories in my simulation. Based on production data in Lisker and Abramson (1964a), I set $\mu_{/p/}$ at 60 ms VOT. I then ran the error minimization procedure in Matlab and determined these optimal fits of the free parameters involved in the identification simulation: $\mu_{/b/} = -0.3\ ms$, $\sigma^2_{/p/} + \sigma^2_S = 96.3$, and $\sigma^2_{/b/} + \sigma^2_S = 336.2$ (See Table 3.5 for comparison). The choice of which mean to set was arbitrary since the simulation is symmetric, meaning that if I set the other mean to the resulting value found in the simulation and re-ran the whole simulation, then I would get 60 ms as the mean for the /p/ category. Of course, this becomes relevant if the mean that we find for the other category does not correspond to what we would expect based on production data. Additionally, the choice of which mean we set becomes more important when we don't have a

[3]The central measure of discriminability that Wood (1976) reported was $-ln(\eta)$, which is monotonically related to the d' parameter. They also included a unit square graph that showed the relationship between hit rates and false alarm rates. By measuring these distances by hand I was able to recover the values needed to compute d' scores. I then used these d' scores in my simulations to keep the measures of perceptual distance consistent with that used for Fricatives. In computing the d' scores I used the standard z-transform independent-observation model, which is the standard in speech perception literature. I did not apply the correction for roving experiments using the differencing model for AX tasks discussed in Macmillan and Creelman (2005)

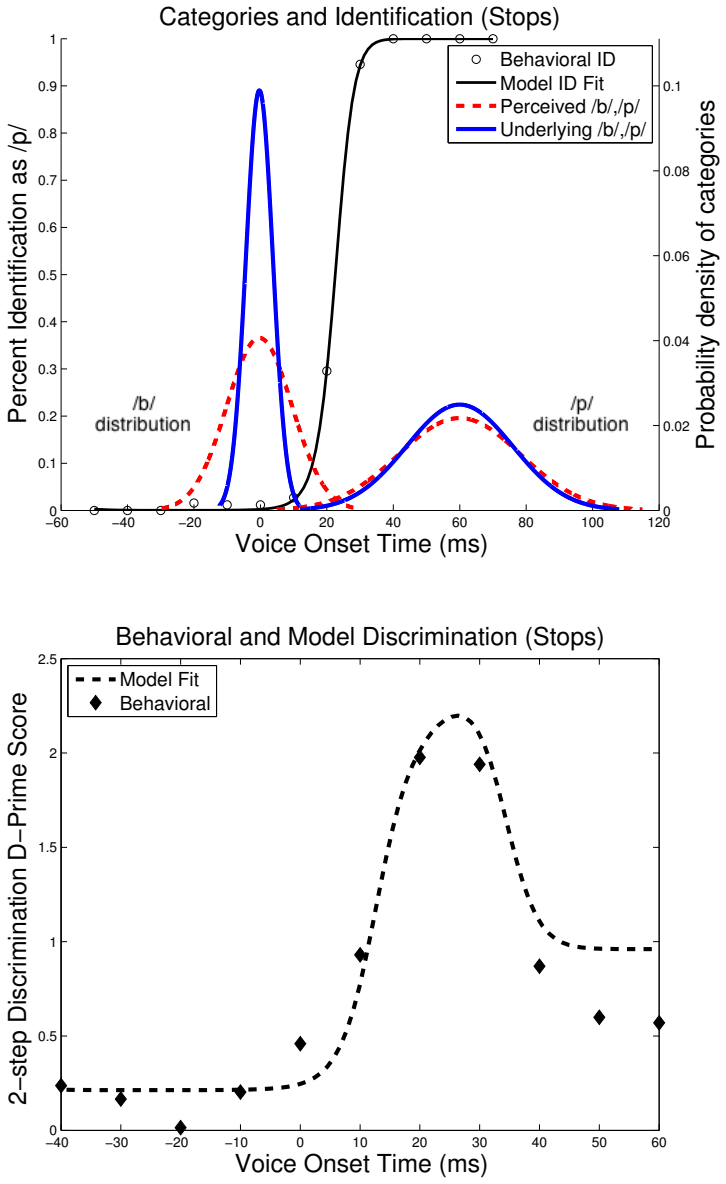


Figure 3.3: Identification and Discrimination fitting for Stop Consonants: (a) Identification fitting, perceived, and underlying categories for stop consonant simulations along the /b/-/b/ continuum, (b) d' scores in bar graph with model fit for discrimination data in black diamonds

priori knowledge of expected category centers. In the case of voice onset times, we can verify that the found value fits with known production data. However, there may be cases where one mean is known and the other is not. Or, as in this case, one of the means has a different ontological status from the other. A VOT around zero corresponds with temporal co-incidence. There is something special about this which leads all speakers of standard English to have the same mean for the voiced category. However, the mean value for the voiceless category is arbitrary, based on experience, and varies cross-linguistically. Even within the same English dialect, speakers may have different values for it, so if we use the model to fit individual speakers to see if they share similar representations for the voiceless category, it would be important to start by setting the mean for the voiced category. I will come back to the question of individual differences in the discussion in section 5.5.

The fit of the model to identification data can be seen in Figure 3.3a. In the same figure, we also see the category structure of the perceived categories, reflecting the two means as well as sums of variances for each category. Lastly, we can look at the inferred second mean for the /b/ category. The value of -0.3 ms is very close to the mode of production data found in Lisker and Abramson (1964a), which I present in figure 3.4. Almost all productions for /b/ were located at zero in that study, though there were some rare exceptions with extreme pre-voicing that they found, which could be due to dialectal differences.

Fitting the discrimination data we get the following value for the individual variances in our model fit: $\sigma^2_{/p/} = 256.2$, $\sigma^2_{/b/} = 16.3$, and $\sigma^2_S = 80.0$ The physical fit of the model to the discrimination data can be seen in Figure 3.3b. In addition

to providing a good overall fit, the model is able to accurately predict the lower within-category discriminability of voiced stops relative to voiceless stops. This can be seen in Figure 3.3b, where the left side of the distribution is substantially lower than the tail on the right.

The underlying categories can be seen, overlaid over the identification data and perceived categories, in Figure 3.3a. Before considering the warping ratio for stop consonants it is interesting to consider how well the underlying categories correspond to those in actual human production. In Figure 3.4 I put the model categories reflecting the inferred parameters for $\mu_{/b/}$, $\mu_{/p}$, $\sigma^2_{/b/}$, and $\sigma^2_{/p/}$, on top of the the distributions found in production studies by Lisker and Abramson (1964a). We see that the distribution of the models is almost perfectly aligned, further lending credence to the findings here.

Based on the parameters found for stop consonants, we get the following values for the critical τ ratios:

- $\tau_{/p/} = 3.12$
- $\tau_{/b/} = 0.20$

As would be expected based on previous findings, the critical variance ratio for stop consonants is found to be substantially lower than that of vowels for both the voiced and voiceless categories. In terms of the model, this means that stop consonants have less meaningful within-category variance and more noise in the signal. This leads the listener to rely to a greater degree to underlying knowledge of categories, leading to greater pull toward category centers, and in turn to what we

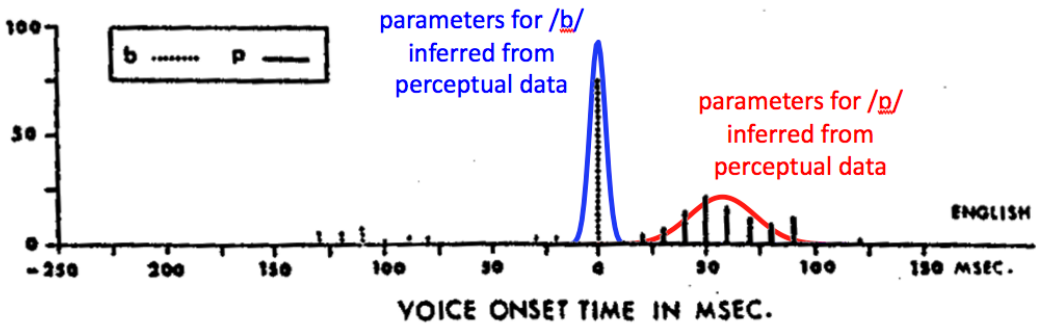


Figure 3.4: Production data for the /b/-/p/ continuum (Reprinted from Lisker and Abramson (1964a)) overlaid with underlying categories found by the model

know as categorical perception. This perceptual bias is especially pronounced for the voiced stop consonant, /b/, due to the low variance in the underlying category.

The corresponding warping views for these values can be seen in in context of other phonetic categories in figure 3.6b for the /p/ category and figure 3.6c for the /b/ category. We can see from these figures just how pronounced this pull toward category centers is, especially compared to that of vowels.

3.4.3.2 Fricative Identification and Discrimination Data Collection

Categorical effects with fricatives vary, with some research findings close to categorical perception, while other found near continuous perception closer to that of vowels. Based on these results, in the current model, one would expect to find that the relationship between meaningful category variance and noise variance is somewhere in between that of stop consonants and vowels. Fricatives, then, provide an important data point because they not only show that the model extends beyond the commonly studied cases of vowels and stop consonants, but gives further insight

into the nature of the relationship between behavioral results and the τ ratio. This continues to build the case for moving past a separate explanation of categorical effects in perception of different phonemes. However, because of past studies employing different paradigms and arriving at different conclusions with fricatives, I first need a reliable data set of behavioral performance for fricative perception. I focus on the continuum connecting the sibilant fricatives /s/ and /ʃ/. As we shift from one to another, we are changing the mean spectral locations used for identifying the fricatives, creating a one-dimensional continuum akin to the voice onset timing in stop consonants or moving through F1-F2 space for vowels. Therefore, this continuum fits well into the existing simulation framework. In this section I describe the behavioral experiments conducted on this continuum along with colleagues at University of Maryland. Further details of the work, along with related MEG experiments, are reported in Lago et al. (2010) and Lago, Scharinger, Kronrod, and Idsardi (Submitted).

The /s/-/ʃ/ continuum was created by varying the two central frication frequencies (spectral peaks, F5 & F6) of the noise portion of a consonant-vowel syllable with a standard /i/ vowel. This continuum relates to the shift in place of articulation moving from alveolar to palatal. The stimuli were artificially created based on human productions, by employing a procedure similar to McQueen, Jesse, and Norris (2009). Utilizing the Klattworks interface (McMurray, 2009), frication, aspiration, voicing, formant bandwidths, and amplitudes were set by hand and smooth logistics were applied to create the initial stimuli. The results were then converted to formats readable by Praat (Boersma & Weenink, 2005), shifting the relevant

Stimulus (S_i)	**Barks**		**Frequencies**		% Ident. of S_i as /s/	$[S_{i-1}, S_{i+1}]$ Disc. d'
	F_5	F_6	F_5	F_6		
S_0	14.5	15.5	2501	2915	0.017	
S_1	15.0	16.0	2698	3152	0.033	0.15
S_2	15.5	16.5	2915	3413	0.000	0.84
S_3	16.0	17.0	3152	3702	0.033	0.63
S_4	16.5	17.5	3413	4025	0.058	1.40
S_5	17.0	18.0	3702	4386	0.133	1.08
S_6	17.5	18.5	4025	4794	0.525	1.50
S_7	18.0	19.0	4386	5258	0.825	1.27
S_8	18.5	19.5	4794	5790	0.992	1.35
S_9	19.0	20.0	5258	6407	0.967	0.91
S_{10}	19.5	20.5	5790	7131	0.992	

Table 3.4: Central Frication Frequencies (in Barks and Hertz, labeled as F5 and F6) for stimuli used in the behavioral experiments by Lago et al. (2010). Identification study results shown as percent of time stimulus identified as the phoneme /s/. Discrimination results are d' scores for two step discrimination, where the score on line i is the discrimination score for the pair of stimuli on lines i-1 and i+1 (e.g. 0.6301481 on line S_3 is discrimination score for S_2 and S_4).

frication frequencies along the way. The values for the continuum are presented in Table 3.4 in the first three columns, with values reported in Barks, a psychologically plausible acoustic scale (Zwicker, 1961). The synthesized stimuli consisted of a 210 ms fricative portion, followed by the vowel /i/ with a duration of 270 ms resulting in the words [ʃi] ('she') and [si] ('see'). The synthesized vowel had formant values indicative of a male, set to 300, 2020, 2960, and 3300 for F1, F2, F3, and F4, respectively.

The participants in the experiment were eleven undergraduate students (9 females, age range = 19-31 years, mean age = 21.8) who completed the study for course credit. All were native English speakers, but we did not check whether they were also fluent in any other languages. All were strongly right handed (Oldfield, 1971) with no history of auditory disfunction. Each participant conducted an identification task, an AX discrimination task, and a goodness judgment task. Identification

and discrimination were both forced choice button press and the goodness task employed a judgment on how well a certain stimuli represented a specific category, either /ʃ/ or /s/, on a 5-point Likert scale. In the identification task, participants heard sounds from a 10-step [ʃi]-[si] continuum and were instructed to classify them as either [ʃi] or [si] by means of button presses. Response-button assignment was counter-balanced across participants. Each trial began with a 500 ms blank screen. Then a fixation cross appeared for 300 ms and was immediately followed by a sound from the continuum. Each sound was repeated ten times across the experiment in a randomized order. If a button press was not registered 4 seconds after trial onset, the next trial began.

In the discrimination task, participants gave same/different judgments about sound pairs taken from the same continuum. The acoustic distance varied between the sounds in each pair, which could be 2 (e.g. stimuli 1-3), 4 (e.g. 1-5), 6 (e.g. 2-8), 8 (e.g. 1-9) or 10 steps apart[4]. There were 94 pairs and each pair was repeated 3 times across the experiment. Same/different trials were presented in a 46:54 ratio. Each trial started with a fixation cross that was displayed for 300 ms. Immediately after, the sounds were presented, separated by a 300 ms inter-stimulus interval (ISI). The next trial began when participants provided a judgment, or if no response was recorded after 2 seconds. Trials were separated by 500 ms during which the screen remained blank.

In the identification task, identification scores were calculated as the mean

[4]Only the 2-step discrimination results are reported directly here and used in the model simulation for consistency with the other phonemes examined in this dissertation

of participants responses for each step in the continuum. In the discrimination task, we determined perceptual sensitivity by calculating d' scores for each sound pair (Macmillan & Creelman, 2005). The log-linear correction method described in (Hautus, 1995) was used to avoid the appearance of non-finite values in the case of extreme false alarm or hit rates. However, we did not use an additional correction for roving experiments using the differencing model for AX discrimination experiments discussed in Macmillan and Creelman (2005), instead settling on the standard z-transform independent-observation model, which is more standard in speech perception literature. The results of the identification and discrimination task are shown in Table 3.4 in the final two columns. The results showed fairly categorical perception of the stimuli, though with perhaps a shallower discrimination peak that one might expect, suggesting representation that incorporates some more of the acoustic signal beyond pure category assignment.

3.4.3.3 Simulation 2: Fricatives

In this section I report the simulation findings from fitting the model to the behavioral data we collected in a set of experiments described in the previous section. Because the fricatives were synthesized using two frequencies, it was necessary to pick a single measure to represent the continuum. Since the two formants were always 1 Bark step apart, it did not matter whether we chose one, the other, or a central measure. Hence, I just picked the F5 measure as the representative one for all simulations. As before, I need to set one of the two category means before fitting

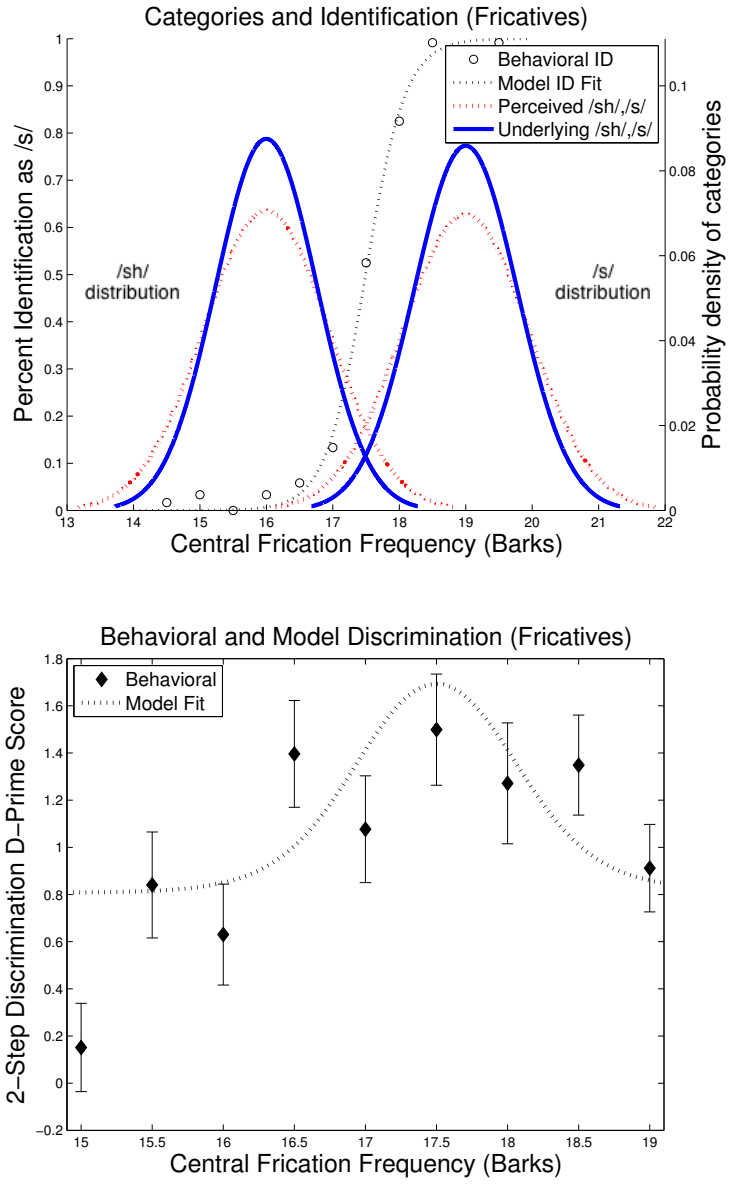


Figure 3.5: Identification and Discrimination fitting for Fricatives: (a) Identification fitting, perceived, and underlying categories for fricatives simulations along the /ʃ/-/s/ continuum, (b) d' scores with error bars in bar graph with model fit for discrimination data in black diamonds

the behavioral data. Based on productions by an adult male participant, I set the value of $\mu_{/s/}$ to be 19.0 Barks. I then proceeded to fit the first set of parameters via an error minimization procedure in Matlab. The optimal model fit produces the following parameters: $\mu_{/ʃ/} = 15.99$ Barks, $\sigma^2_{/ʃ/} + \sigma^2_S = 0.909$, and $\sigma^2_{/s/} + \sigma^2_S = 0.887$ (See Table 3.5 for comparison). The fit of the model to the underlying data can be see in the identification curve overlaid over the behavioral data points in Figure 3.5a. In the same figure we also see the perceived categories, representing the parameters for the means of the categories as well as the sums of meaningful and noise variance.

Fitting the discrimination data we get the following values for the individual variances in our model fit: $\sigma^2_{/ʃ/} = 0.5992$, $\sigma^2_{/s/} = 0.5772$, and $\sigma^2_S = 0.3098$ (See Table 3.5 for comparison). The fit of the model to the discrimination data can be seen in Figure 3.5b. Although the fit here is not as good as that of the stop consonants, note that the peak in discrimination in both the underlying data and the model fit occurs in the same location, and that location corresponds to the inflection point of the identification curve. While the model does not fit every mean of the Participants' d' data, if we look at the variability between participants represented by the error bars, we see that the model fits the data within the margin of error for almost all data points. It is possible that fitting a separate model to each participant would give us insight into individual variability and the source of this disparity, but the data would be sparse, and this is past the scope of the present work.

Based on the parameters found for fricatives , we get the following values for the critical τ ratios:

- $\tau_{/\int/} = 1.86$
- $\tau_{/s/} = 1.93$

As mentioned already, previous studies have been mixed as to the degree of categorical effects for fricatives. Therefore, I would expect values for the ratio somewhere between the extremes for categorical perception and continuous perception. I find this to a limited extent. We can see how these values compare to the others in Figure 3.7. The fricatives are much further down the continuum than the vowels, and well above the voiced stop consonants, though roughly close to the voiceless stop consonant /p/. Based on the model, this means that the listener is depending on the underlying categories more than for vowels, but also paying more attention to the finer acoustic detail than for voiced stop consonants. However, given that the fricatives are lower on the continuum than the voiceless stops and we have only considered one case of stop consonants and fricatives, we would want to see simulations for other stop consants and fricatives before saying precisely what the relationship is between these two phoneme types. The warping view for the fricatives can be seen in Figure 3.6d in the context of other phonetic categories. Given the proximal τ values found for the two sibilant fricatives, only /s/ is pictured as the representative case.

3.4.3.4 Simulation Summary

The simulations presented in this section have shown that the model both provides good empirical fits to behavioral data and derives ratio values consistent

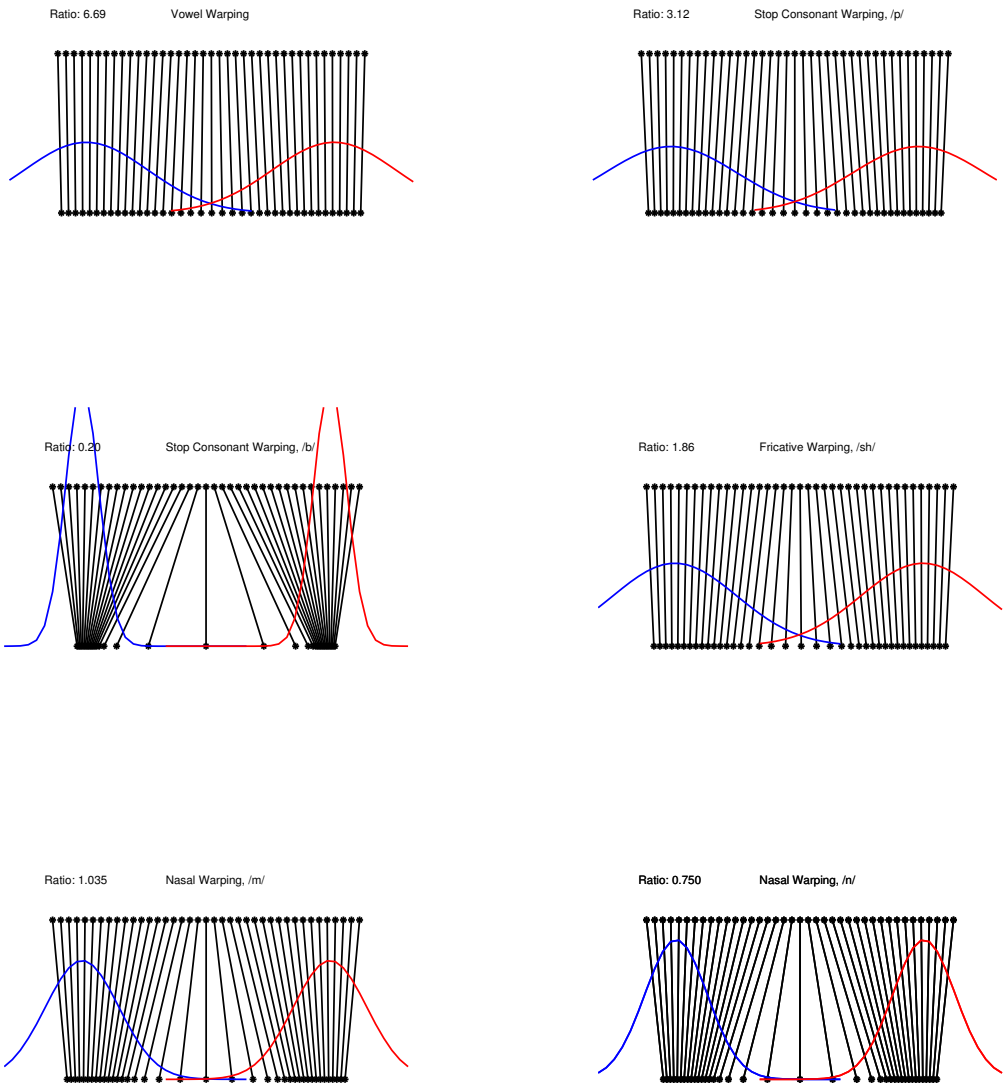


Figure 3.6: Warping Representation for 6 types of phonemes: (a) Vowels, (b) voiceless stop consonant /p/, (c) voiced stop consonant /b/, (d) sibilant fricatives /s/ and /ʃ/, (e) bilabial nasal /m/, and (f) alveolar nasal /n/

Simulation	**Means**		**Variances**			**Meaningful:Noise Variance Ratio (τ)**
	μ_{c_1}	μ_{c_2}	$\sigma^2_{c_1}$	$\sigma^2_{c_2}$	σ^2_S	
Vowels (Equal Variance)	F1=423 Hz F2=1936 Hz	F1=224 Hz F2=2413 Hz	5,873	5,873 (Mels)	878	6.69
Stop Consonants (Unequal Variance)	-0.3 ms	60 ms	16.3	256.2	80.0	/b/: 0.20 , /p/: 3.12
Fricatives (Unequal Variance)	15.99 Barks	19.0 Barks	0.599	0.577	0.310	/ʃ/: 1.86, /s/: 1.93
Nasals (Unequal Variance)	922 Hz	1600 Hz	10,143	7345	9798	/m/: 1.04, /n/: 0.75

Table 3.5: Best fitting model parameters for vowels (N. H. Feldman et al., 2009), stop consonants, fricatives, and nasals.

with the continuum of categorical effect sizes. First, I needed to make sure that the model derives a reasonable set of parameters. We saw in Figures 3.3 through 3.5b that the model is able to provide reasonable fits to identification and discrimination data from multiple sets of experiments. Not only were the means and variances reasonable compared to production data, but further details such as differences in discriminability between voiceless and voiced stop consonants were captured as well. Fricative data was also fit very well. Although discrimination was not a perfect fit, the general dynamics were captured, and between the identification and discrimination, the fit was adequate for making further judgments about the derived parameters. Second, I wanted to see whether the model captures categorical effects with a single qualitative approach and parametric variation within a single model. To do this, I considered the critical ratio of meaningful to noise variance for each phoneme to see if it is in line with predictions set forth by the model and conforms to previous findings. Based on past research, I expected vowels to show a very high

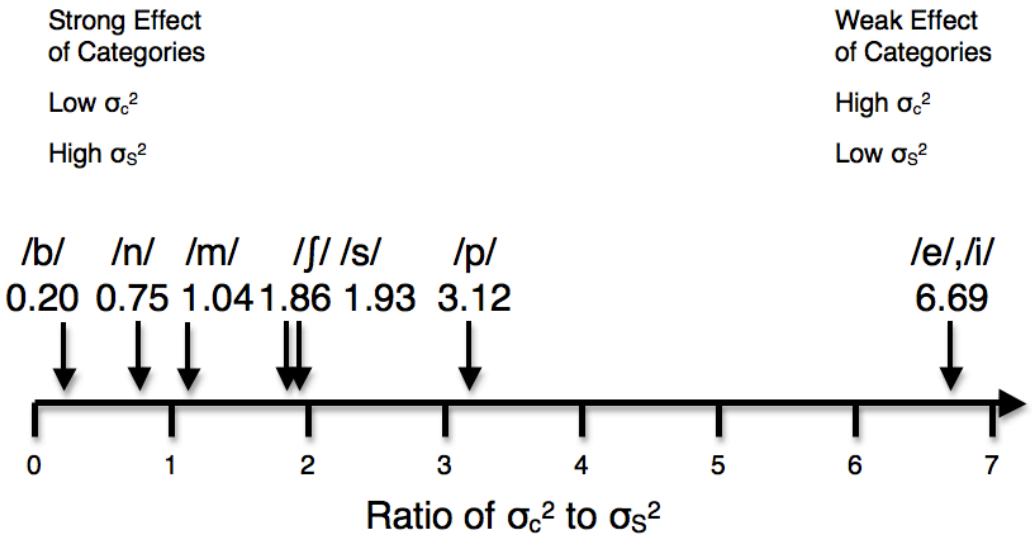


Figure 3.7: Fitted τ values for vowels, stop consonants, fricatives, and nasals

ratio, reflecting weak effects of categories with high meaningful variance and low noise variance. Conversely, I expected stop consonants to show a low ratio, reflecting strong effects of categories with low meaningful variance and high noise variance. Finally, I expected fricatives to be somewhere in between, due to mixed previous findings in perception experiments, and further supported by analyses presented in the section on phoneme differences. Findings were largely found to conform to predictions, with the single exception being the voiceless stop consonant ratio being further along the continuum than either of the sibilant fricatives. Figure 3.7 provides a very succinct visual aid to the relevant continuum of ratio values, while figure 3.6 shows the relevant warping demonstrations for these same ratio values.

3.4.4 Correlates of Degree of Categorical Effects in a Unified Model

Before considering nasals, which represent a novel test of the model as well as requiring a different approach to fitting the discrimination data, I would like to consider the implications of the findings presented thus far. I look at how τ might reflect other characteristics of vowels, stop consonants, and fricatives that have been discussed in the literature. Particularly, what would a possible mapping be between features of the phonemes and where they lie on the τ continuum. And what testable predictions would we be able to make based on this analysis that can shed more light on this model of speech perception?

First, I consider the set of features used to describe the phonemes as well as where the phonemes fall on the sonority hierarchy. The sonority hierarchy is a scale that ranks phonetic segments (or phonemes, depending on context) by their relative resonance in relation to other segments (often simply referred to as amplitude) (Selkirk, 1984). In the original formulation, it looked something like:

/a/>/e,o/>/i,u/>/r/>/l/>/m,n/>/z,v,ð/>/s,f,θ/>/dʒ/>/tʃ/>/b,d,g/>/p,t,k/

Grouping segments of the hierarchy into natural classes, it can be rewritten as:

Vowels > Liquids > Nasals > Fricatives > Affricates > Stop Consonants

The vowels are the most sonorous and as we go from left to right on the ordering the sounds become less sonorous. This continuum corresponds to how loudly these phonetic categories are produced, patterns in syllable structure, rules on co-occurence, and assimilation that is conditioned on locations on the hierarchy or changes in

adjacent sonority. With the exception of voiceless stop consonants falling past the fricatives, this hierarchy captures the pattern observed in the simulations so far, with the degree of categorical effects along the τ continuum closely following the location of the relevant phonemes on the sonority hierarchy. At one end are vowels, that have been shown to have quite continuous perception. At the other, the stop consonants, which exhibit a very strong effect of categories. Fricatives , which have had mixed results, are between the end-points, though before the voiceless stops. The correspondence between the sonority hierarchy and the τ continuum suggests that there might be a relationship between the relative resonance of speech segments and their respective degree of categorical effects. We therefore need to examine other phonemes before we can say whether there is a true continuum/hierarchy or whether we are seeing a coarser division of consonants on one end of the continuum and vowels on the other.

Beyond sonority, it is possible that lower level phonological distinctive features making up the phonemes are what guide the degree of categorical effects of a category. In terms of the model, this would mean that sounds that possess or lack a certain feature might be more or less prone to noise variance. In terms of features, stop consonants and fricatives share certain feature values to the exclusion of vowels, namely both lacking the syllabic, approximant, and sonorant value. Fricatives also share certain feature values with vowels, to the exclusion of stop consonants, namely being positive for the continuant and delayed release features. If having these features insulates a phoneme from noise interference (or, likewise, missing these features exposes a phoneme to noise interference), then we would also predict

the most continuous perception of vowels (positive for all five features), the most noise interference, and hence categorical perception, for stop consonants (negative for all five features), and fricatives somewhere between the end points (positive for 2 features). Hence, breaking down categorical effects along featural lines can possibly help account for the graded degree of categorical effects as we move along the sonority hierarchy and the τ continuum. Here, again, testing a broader set of phonemes can shed light on the question.

Features, however, are an abstract representation of phonemes. I also consider acoustic cues that are necessary and sufficient for identification and discrimination of various phonetic categories. Following the terminology of the research reviewed in this dissertation, I consider cues along both the static-dynamic and the spectral-temporal dimensions. Static properties are ones that can be identified by looking at a small time slice of the spectral makeup of the sound and dynamic ones are ones that refer to change over time in the signal. Spectral properties refer to the frequencies that make up the signal, as seen in a spectrogram, while temporal properties are time locked and depend on relative timing of parts of the signal. Most phonemes are identified by some combination of these types of properties. There has been a sizable body of research on figuring out precisely which sorts of cues are most useful for identification of various phonemes. While there is evidence that all types of cues contribute to the identification and discrimination of all types of phonemes, certain ones have been found to be the most critical, and in many instances, sufficient for the patterns of behavior seen in natural stimuli studies. Here I consider the cues that are specifically relevant to the types of stimuli I consider in the studies in the

dissertation. At least in terms of all the L1 studies, this includes consonants in /CV/ syllables and standalone vowels.

Stop consonants have largely been found to be identified by static temporal cues like VOT for the voicing distinction (Carney, Widin, & Viemeister, 1977) and dynamic spectral cues such as rapid changes in spectra at release (K. N. Stevens & Blumstein, 1978), time-varying spectral features (Kewley-Port, 1983), locus equations (H. A. Sussman, McCaffrey, & Matthews, 1991), and changes in moments (Forrest, Weismer, Milenkovic, & Dougall, 1988). Vowels, at least when they are standalone without surrounding consonants, can be largely identified by the first and second formants, representing the steady state peaks of resonant energy with certain bandwidths, a static spectral cue. However, while my L1 experiments are all based on standalone vowels, it has been shown that whenever examining natural human-produced stimuli, coarticulated vowels would always be identified more accurately than standalone vowels (Strange, 1989). When considering identification of coarticulated vowels in a normal speech stream, various measures of vowel inherent spectral change (VISC) such as formant transitions play a role, changing the cues to identification to also include dynamic spectral cues and increasing reliance on temporal cues such as duration (see J. M. Hillenbrand (2013) for a thorough review of the history of research on vowel perception). I will consider the difference between standalone and coarticulated vowels as well as consonants in Section 5.5. Fricatives have the widest array of critical cues, covering dynamic spectral cues in the form of locus equations (Jongman et al., 2000) and changes in moments (Forrest et al., 1988), static temporal noise duration (Strevens, 1960; Jassem, 1962), and

static spectral cues including F2 onset frequency, spectral moments, and spectral peak location (Jongman et al., 2000). This list represents cues that were found to be strongly statistically significant in allowing a listener to properly perceive a difference or change in the phonemes being played to them. The methodologies of the various studies differed greatly, but all of them examined effects of multiple cues and pulled out the ones that were most significant.

I observe that for identification of consonants in /CV/ syllables and for standalone vowels, there are overlaps in the types of cues used to perceive fricatives with both stop consonants and vowels, while there is no overlap between stop consonants and vowels. If cues such as the static spectral ones are prone to continuous perception, and cues such as dynamic spectral and static temporal ones lead to categorical perception, then the distribution of cues and phonemes corresponds well with previous work as well as the modeling results in this work. Borrowing terminology used in Pisoni (1975), I consider two modes of perception: acoustic and phonetic. Whereas in past work these modes were reserved for a process level account of how a decision in a discrimination experiment was being made, I apply them to the actual processing of the stimuli. If static spectral cues trigger acoustic processing and dynamic spectral and static temporal cues trigger phonetic processing, then we get a pattern similar to that considered above for the featural or sonority account. Further, the perception of different cues may be more or less prone to noise variance, connecting to the computational account of the present model. Particularly, it is likely that static spectral cues, with their longer duration and continuous presence of the signal to be detected, would in fact be less prone to interference. While fur-

ther work would be needed to determine if this is correct, it is possible that certain cues are processed in different ways precisely because of the associated balance of noise to meaningful variance within their perception. Examining a greater variety of phonemes that are identified by different cues would help establish the strength of this correlation. Also, it would allow us to examine which acoustic cues to perception are most strongly correlated with the degree of categorical effects. Of course, things would change when we consider coarticulated vowels and consonants appearing after vowels in a syllable. In Section 5.5 I will consider this further and highlight why this overlap in cues in continuous speech might actually predict in this similar fashion differences in the categorical effects on vowels in continuous vs. standalone production.

I would also like to revisit the notion of restructuring in the speech stream, whereby sounds exhibit varying acoustic realizations as a result of varying context, even though they share a common underlying phonemic form. Liberman et al. (1967) examined how the acoustics of various phonemes varied between contexts. They referred to this variability as the extent to which a phoneme exhibits restructuring. They found that stop consonants show a large amount of restructuring, while steady-state vowels show much less, once speaker normalization and speaking rates are accounted for. In their study, they particularly examined place of articulation for stop consonants, though similar studies showed similar findings for both manner of articulation and voicing as well (Lisker & Abramson, 1964b; Liberman et al., 1954). They proposed that the degree of restructuring correlated with the categorical effects on perception, namely that those sounds that exhibited a large

amount of restructuring were perceived categorically and those that did not are perceived continuously. This restructuring can be viewed as a form of variability that is noisy because it does not allow for a direct mapping between the acoustic cues and the underlying phoneme to be perceived. Hence, per our model, this noise pushes the listener to use their underlying knowledge of the categories more in order to handle the variable input, leading to greater categorical effects for stop consonants than for vowels. By this particular metric, fricatives were found to pattern more with vowels, showing little restructuring (Harris, 1958; Hughes & Halle, 1956). This suggests that restructuring can't be the driving force behind this pattern since we found fricatives to be closer to the stop consonant end of the continuum in my simulations.

3.4.4.1 Verifying Correlates

Thus far, I have shown that the unified Bayesian model accounts for stop consonant, fricative, and vowel perception, but only for a specific set of phonemes, namely bilabial stop consonants, sibilant fricatives, and front non-low tense vowels. While this does show that we can account for a range of classic behavioral findings in a single model, this does leave many phonetic categories to be addressed. While one further step could be to apply the model to as many phonemes as possible, a better focal point is to address the possible correlation between these categorical effects and the related perceptual cues discussed above such as sonority or phonetic features. In particular, a logical next step is to apply this model to behavioral

data from nasal and liquid perception, as this would allow us to see the extent of similarities between the τ continuum and the sonority hierarchy. For nasals, we can vary the initial formant transitions to create a place of articulation continuum from /m/ to /n/. Not only would this add another phonetic category, but it would also show that the model works for place of articulation dimensions, which have not been thus far considered. For liquids, there are two continua that can be investigated. A spectral cue of either frequency onset or F2/F3 transitions have been shown to guide identification and discrimination. A temporal cue of either relative duration of initial steady state or F1 transition is also significant. They were shown to be in a trading relation to each other in tests of identification and discrimination of American liquids (Polka & Strange, 1985).

The case of the liquids, however, warrants a consideration of the number of dimensions that the model handles. It is a simplification to treat any continua of sounds as purely varying in a single dimension. While it is a reasonable assumption for things like voice onset time variation and central frication frequencies, other sounds can not be as easily studied with a single linear variation. Even for vowels, if we wanted to consider a case with more than two phonemes it would be impossible to have them lie along a single line. Correspondingly, the model can be expanded to account for multiple dimensions. Treatment of the multiple dimensions would depend on how they interact with each other. Adapting terms from Garner (1974), we can refer to dimension as being either separable or integral. Integral dimensions are ones that are not necessarily processed independently and integrated in the perceptual system. Separable dimensions are ones for which a subject can perform

operations over one dimension without regard for the other one. In psychology studies, integral dimensions are typically combined so that distance can be described as a Euclidean metric, while for separable dimensions we tend to employ a city-block metric (Johannesson & Lundagard, 2001). With vowel perception, F1 and F2 perception is integral to each other, and we can therefore employ the Gaussian distance metric to describe distance between two vowels in this space. However spectral and temporal dimensions are integral to each other, with experiments confirming that dimensions like frequency and durations do not interact decisionally (Silbert, Townsend, & Lentz, 2009). Hence, it is likely that the spectral and temporal dimensions for liquid perception would be integral.

In general, the model can be expanded to account for effects in perception for an arbitrary number of classes of sounds and arbitrary number of dimensions. This would also allow us to study more ecologically valid sound perception in situations where identification is not constrained to a binary choice and multiple sounds affect the representations that underlie discrimination ability. Of course this introduces new considerations as we may have to combine both integral and separable dimensions. Past research suggests that in these cases it is best to treat each dimension separately and then combine them additively rather than employing any typical metric such as Gaussian distance of the city-block measure (Johannesson & Lundagard, 2001). An initial version of a multidimensional model that could be used for integral dimensions was developed by Barrios (2013). It remains for future work to examine working with separable or more than 2 dimensions.

3.4.5 Nasal Consonants and Monte Carlo Simulations

As a final test of the model in native language perception included in this dissertation, as well as to develop a more in-depth understanding of the distribution of τ values, I now turn to nasal phonemes. Particularly, I examine the place of articulation continuum between /m/ and /n/. There are several reasons that this constitutes an important next step. First, this allows us to explore a new manner of phoneme (nasal stops) as well as a novel continuum (place of articulation as correlated with formant onset transitions). Second, this can help settle the proposal concerning the relationship between the sonority hierarchy and degree of categorical effects. Third, we can explore the relationship of specific features to the τ continuum and see how multiple cues relate to degree of categorical effects in perception when they have conflicting predictions; Nasal stop consonants contain strong steady nasal resonances but the place of articulation continuum is correlated to the same cue as for oral stops, so they do not clearly align with the other phonemes I have considered thus far. Because the continuum is along a single dimension, I can largely use the same fitting procedure as before. However, the best behavioral data available for the nasal continuum does not provide d' scores for discrimination, but rather contains accuracy scores. Hence, we need a different way of fitting the model to the data. In this section I briefly describe my approach to fitting this accuracy data and then go on to describe the model simulation as for the other phonemes.

3.4.5.1 Monte Carlo Simulation for Accuracy Data

In the previous simulations, I had access to d' data, which is a measure of perceptual distance. I hypothesized that this distance would be directly proportional to actual distance between inferred target productions, making the fitting of the model to the discrimination data a matter of finding an optimal noise variance and a constant to scale the d' data to the target production values. When we instead have accuracy data, there is no direct relationship to perceptual distance. In particular, this is because we do not know how the Participants behaved on 'Same' trials at different points in the continuum. Hence, I need to have a way of relating any given setting of means and variances in the model to a particular accuracy score. The way I do this is via a Monte Carlo simulation. Given a set of parameters in the model, I derive a posterior distribution for target productions of stimuli. Then I sample target productions from these posterior distributions and make decisions based on these targets in the discrimination task. The particulars change depending on the discrimination task. For an AX task the 'Same'-'Different' decision is based on an ϵ distance, where if the two target productions are closer to each other than distance ϵ they are called 'Same', otherwise 'Different'. For an ABX or Triad paradigm the response is based on the relative distances between the three stimuli. In an ABX task, we compute the distances $\Delta_{A,X}$ and $\Delta_{B,X}$, Then if $\Delta_{A,X} < \Delta_{B,X}$ we say that X is the same as A. If $\Delta_{B,X} < \Delta_{A,X}$, we say that X is the same as B. If the distances are the same a coin is flipped. Finally, in the Triad paradigm, the goal is to choose the odd man out. If we call the three

Stimulus (S_i)	F2 (Hz)	F3 (Hz)	% Ident. of S_i as /n/	$[S_{i-1}, S_{i+1}]$ Disc. Acc.
S_1	921	2018	0.020	
S_1	1075	2180	0.020	0.45
S_2	1155	2348	0.020	0.51
S_3	1232	2525	0.180	0.73
S_4	1312	2694	0.820	0.75
S_5	1386	2862	0.940	0.57
S_6	1465	3026	0.990	0.48
S_7	1541	3195	1.000	0.50
S_9	1695	3363	1.000	

Table 3.6: Formant transition values for stimuli used in the behavioral experiments by Miller and Eimas (1977). Identification study results shown as percent of time stimulus identified as the phoneme /n/. Discrimination results represent percent accuracy scores for two step discrimination, where the score on line i is the discrimination accuracy for the pair of stimuli on lines i-1 and i+1 (e.g. 0.73 on line S_3 is discrimination score for S_2 and S_4).

stimuli A, B, and C, we compute the three distances $\Delta_{A,B}$, $\Delta_{B,C}$, and $\Delta_{C,A}$. Then we choose the smallest distance and say that those two stimuli are the same ones with the third one being the odd man out and hence 'Different'. As before, cases of ties are decided with a coin flip. When these simulations are run for enough of iterations, the result is an empirical accuracy score, or the percentage of trials on which the chosen sampled stimulus actually belonged to the correct posterior distribution. Once these simulated accuracy scores are available for each step in the continuum, we can compute error scores for this set of parameters using the same procedure as I would for d' scores.

3.4.5.2 Simulation 3: Nasals

For simulations with nasals I consider identification and discrimination data from Miller and Eimas (1977). They conducted experiments considering stimuli along two continua: nasal duration and F2/F3 formant transitions. Here I use data

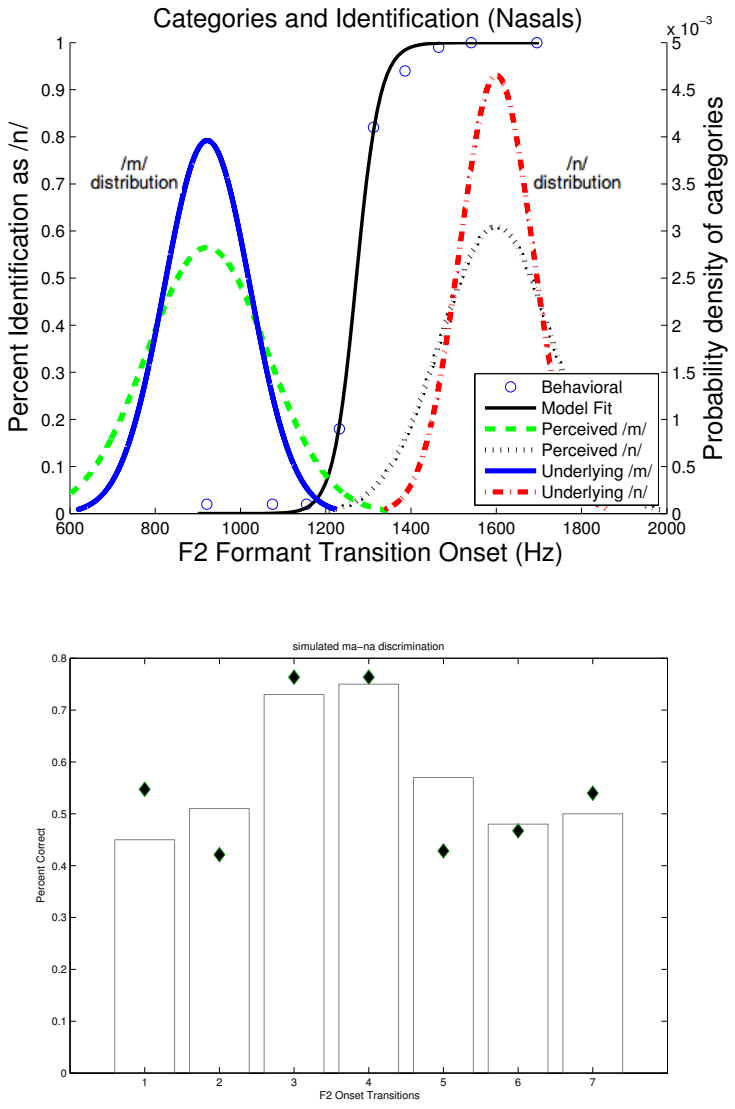


Figure 3.8: Identification and Discrimination fitting for Nasals: (a) Identification fitting, perceived, and underlying categories for nasals simulations along the /ma/-/na/ continuum, (b) Accuracy scores in bar graph with model fit for discrimination data in black diamonds.

from their experiment focusing on the percpeption of /m/ and /n/, as viewed along a continuum of varying initial F2 and F3 values for the formant transition portion of a CV syllable with vowel /a/ (i.e.: /ma/-/na/). Stimulus values can be seen in columns 2 and 3 of Table 3.6. It worth noting that they did not use equally spaced stimuli in their experiment. This allowed them to test with greater sensitivity near the boundary, which requires small step sizes, have the continuum go all the way out to the ends of the categories, which requires a large stimuli range, while still maintain a small number of experimental stimuli. Hence they ended up with larger step sizes near the edges of the continuum than at the center, and as a result, great discrimination scores for those pairs. While this is not a problem for the model since the distance is taken into account when computing perceptual distance, it is important to keep in mind when looking at the graphs of the results and model fit, since we see discrimination going up at both ends of the continuum.

Their identification task was a classic forced-choice identification task. For discrimination, they administered a triad design, where each pair of stimuli 2 steps apart were combined in all possible combinations with two of one and one of the other (i.e. for every stimuli A, B they presented triads AAB, ABA, BAA, ABB, BAB, and BBA). For every presented triad, the participant had to identify which of the three stimuli was the odd man out. I fit the model to their accuracy data, utilizing the simulation presented in the previous section. The values for identification and discrimination results from their experiment can be found in the last two columns of Table 3.6.

As before, I need to set one of the two category means before fitting the

behavioral data. Based on productions by two adult male participants, I set the value of $\mu_{/n/}$ to be 1600 Hz. I then proceeded to fit the first set of parameters via an error minimization procedure in Matlab. The optimal model fit produces the following parameters: $\mu_{/\mathrm{m}/} = 923$ Hz, $\sigma^2_{/\mathrm{m}/} + \sigma^2_S = 1994$, and $\sigma^2_{/n/} + \sigma^2_S = 1714$ (See Table 3.5 for comparison). The fit of the model to the underlying data can be seen in the identification curve overlaid over the behavioral data points in Figure 3.8a. In the same figure we also see the perceived categories, representing the parameters for the means of the categories as well as the sums of meaningful and noise variance.

Fitting the discrimination data we get the following value for the individual variances in our model fit: $\sigma^2_{/m/} = 10,143$, $\sigma^2_{/n/} = 7345.9$, and $\sigma^2_S = 9797.6$ (see Table 3.5 for comparison). The fit of the model to the discrimination data can be seen in Figure 3.8b. Note that the peak in discrimination in both the underlying data and the model fit occurs in the same location, and that location corresponds to the inflection point of the identification curve.

Based on the parameters found for nasals, we get the following values for the critical τ ratios:

- $\tau_{/m/} = 1.035$
- $\tau_{/n/} = 0.750$

The corresponding warping view for these ratios can be seen in Figures 3.6e and 3.6f in the context of other phonetic categories. As mentioned already, previous studies have largely stated that perception for nasals resembled that of categorical perception. Therefore, it is no surprise that the values for the ratio we find are close

to the purely categorical end of the continuum. We can see how these values compare to the others in Figure 3.7. The nasals are right between the voiced stop consonants and the fricatives. This further makes sense since they share some properties with both of them. Like fricatives, they contain periodic noise that can help identify the specific consonant. LIke /b/, they are stop consonant and are voiced. Further, the formant transitions that they are identified by are also used by identify their oral stop consonant counterparts /b/ and /d/. Of course, I did not examine these counterparts, so the cues were different represented in the tau continuum, but the clustering suggests that there might be something about stop consonants that makes them behave a certain way. It would be interesting to examine this further by considering whether all oral and nasal voiced stop consonants fall close to each other on the continuum independently of the particular cue used to identify members along the continuum.

This further makes sense when we consider that they have the same formant transitions as stop consonants, are voiced like the voiced stop consonant here, but contain periodic noise also that can help identify the specific consonant, much akin to the fricatives. The nasals falling right in the middle of the apparent consonant cluster further lends credence to the idea that there is a large divide between the vowels and the consonants and that the consonants form a cluster and not just end up spread out evenly along the continuum.

3.4.5.3 Revised Simulation Analysis

Based on the results from the nasal simulations, we can update our understanding of the relationship between phoneme properties and the τ continuum. First, I will address the relationship of the sonority hierarchy to the values along the continuum. Based on vowel, fricative, and stop consonant results, it was postulated that the τ continuum largely re-capitulated the sonority hierarchy. Based on this prediction, I would have expected the nasal stop consonants to fall squarely between the other phonemes and vowels, as they are more sonorous than stop consonants and fricatives. This same result would also have been predicted from the fact that nasal consonant have a fairly steady, if aperiodic, noise component (the nasal murmur) that adults have been found to reliably use to identify nasals along this continuum (Mayo & Turk, 2005). However, the τ value for the simulation for the /m/-/n/ continuum fell on the other end of the continuum, with only voiced oral stop consonants closer to the purely categorical end of the continuum. This suggests that the sonority hierarchy is unlikely to have any direct relationship to the degree of categorical effects in phoneme perception. More importantly, this suggests that there is something fundamental about voiced stop consonants, whether they are oral or nasal in their manner of articulation, that leads to strong categorical effects in how they are processed[5].

Another possibility put forward had to do with the relationship between fea-

[5]It is possible that the proper characterization is not that of strong categorical effects but rather a lack of strong attention to acoustic detail. While a subtle point, it is a very different claim in regards to which is the default processing mode and which is the result of certain properties of the phonemes.

tures of the phonemes and their degree of categorical processing. Vowels, which are positive for all five features considered (+syllabic, +approximant, +sonorant, +continuant, +delayed release), were perceived the most continuously, suggesting that the presence of any positive feature leads to more continuous perception (or, rather, having a negative value may lead to more categorical perception). The fricatives were found to be more continuous than at least voiced stop consonants, suggesting that being + continuant and + delayed release might have caused this. Meanwhile, both fricatives and stop consonants shared the values of -syllabic, -approximant, and -sonorant. This suggested that due to these three feature values they were clustered together at the categorical end of the continuum. In terms of these five features, nasal stop consonants are -syllabic, -approximant, +sonorant, -continuant, but are not defined for the delayed release feature (Hayes, 2009). Since they share the sonorant feature with vowels, this feature probably does not play a strong role in determining categorical effects, seeing as nasals were found to be strongly categorical. Further, since the stop consonants technically straddled the fricatives, while they both share the same five features, it is unlikely that the continuant or delayed release features play a strong role either. This leaves the features syllabic and approximant that distinguish the consonant cluster from the vowels. If individual features do determine overall degree of categorical effects, it would seem that the combination of -syllabic and -approximant lead to stronger categorical effects. Ideally we would be able to examine precisely how these features are related to the tau continuum and in turn to the degree of categorical effects, examining what about the acoustic correlates of these features leads the listener to greater bias to category

centers. However, this is beyond the scope of this dissertation and a subject for future study. Further, if we wanted to confirm this feature hypothesis, we would want to examine a wider range of phonemes that have other combinations of these features, though this is also left as a line of inquiry for the future.

Another way of grouping phonemes together in regards to how they are processed concerns the cues that are used to identify them. Based on prior simulation analysis, I proposed that static spectral cues may be prone to continuous perception while dynamic spectral and static temporal cues may be more prone to categorical perception. In terms of terminology proposed by Pisoni, static spectral cues might trigger an acoustic processing mode while dynamic spectral and static temporal cues might trigger a phonetic processing mode. In turn, these different cues may be more or less prone to noise interference in their detection. This would provide a direct link to the framework of this computational model. Based on work by Mayo and Turk (2005), nasal stop consonants may be identified by both the nasal murmur (or resonance) at the beginning of the sound, which is a static spectral cue, and by the formant transitions, which are a dynamic spectral cue. Depending on which cue is stronger, nasals might be perceived more or less categorically. Given my findings, I would predict that the transitions constitute a much stronger cue than the nasal murmurs. This is consistent with findings by Mayo and Turk (2005) that showed that adults only used nasal murmur for identification in the presence of hard to distinguish transitions. Further, they found that children had a very hard time using this secondary cue at all, suggesting that it is a secondary in importance and possibly is not as robustly represented as the transitions. Together, this suggests

that individual cues weighted together have the capacity to explain the distribution of τ values we have seen here.

3.4.6 Discussion of Model Applied to L1 Behavioral Evidence

There have classically been various phenomena used to describe speech perception of vowels and consonants. While these phenomena, and associated models, differ in their implementation and make varying assumptions, they describe behavior that has a common basis across phoneme types. This led me to ask whether there is a model of speech perception that can capture all the results under a qualitatively common explanation. I claim that the Bayesian model described in this dissertation is precisely such a model. I have shown that this model is able to capture the behavioral findings for vowels, fricatives, oral stop consonants, and nasal stop consonants. In terms of the features that determine continua which I have explored, I have shown that this model can capture behavioral results along an F1/F2 formant continuum in two-dimensional frequency space, the Voice Onset Time (VOT) continuum measured in milliseconds, central frication frequencies as measured in Bark space, and the place of articulation as measured by formant onset transitions.

There is a principle, or heuristic, often employed in scientific arguments called Ockham's Razor that states that all things being equal, the simplest explanation is the right one. In general, models that make fewer assumptions and limit the number of entities involved in the explanation tend to be preferred over more complex models (see Thorburn (1915, 1918) for a thorough history and various formulations of this

argument). In the spirit of Ockham's Razor I ask whether it is more likely that there are separate underlying phenomena or that the results described for perception of various phonemes are derived from a single source. My work addresses this point directly. I considered behavioral results that were the motivating force for proposals of CP and the PME and showed that the present Bayesian model fits the data from all cases and that they fall neatly along a parametric continuum. This makes the explanation more parsimonious without loss of explanatory power. I therefore put this model forward as a more likely candidate for the behavioral results.

The model captures the behavioral results and provides an explanation for the varying strength of effects by means of considering two sources of variance, meaningful category variance and articulatory and perceptual noise variance. However, the model does not provide an explanation for why certain phonemes would be more prone to noise variance or lead people to pay more attention to the underlying category variance. Nor have I been able thus far to pin down the relationship of this variance ratio to other aspects of the phonemes or the cues known to be used in their perception. There are a couple relationships that I have shown not to hold. First, I considered sound restructuring in the speech stream as a possible source of varying results. This would make sense since restructuring is potentially a form of noise variance since it involves unpredicted changes in surface presentation. However, by this property we would have expected to see fricatives closer to vowels in their degree of categorical effects (i.e. we would have expected more continuous perception of fricatives). Therefore, I do not consider this to be a possible source of the varying effects. I also considered the sonority hierarchy as it relates to the

degree of categorical effects on perception. This is potentially a good fit in the sense that increased sonority could lead the listener to pay more attention to the acoustic cues in the speech signal. However, the only correlation between this hierarchy and the distribution of τ values for the examined phonemes is that vowels fall on one end and voiced stop consonants on the other. Fricatives and nasals show that this relationship does not hold up. In fact, while there is certainly different degrees of category effects among the consonants, there definitely seems to be a large split between vowels and consonants as a group. This suggests the possibility that degrees of categorical effects have to do with the feature makeup of a phoneme or phoneme continuum. Since all the consonants share the -syllabic and -approximant feature values, the detection of this feature might lead to strong categorical effects. Within the consonants, other features may play a role in distinguishing very strong from just modest effects of categories.

Digging deeper, I suggest that the perception of multiple individual cues together determine the degree of categorical effects, but they are weighed differently. The idea of cues being weighted in order to arrive at a category determination is not new. Many models have been proposed to show how people integrate cues when perceiving speech. Two fairly classic models include the Fuzzy Logical Model of Perception (Massaro & Oden, 1980; Oden & Massaro, 1978) and the Normal A Posteriori Probability (NAPP) model (Nearey, 1997; Nearey & Assmann, 1986; Nearey & Hogan, 1986). The learning of cue weights has also been treated together with category acquisition as an unsupervised clustering process (McMurray, Aslin, & Toscano, 2009; Maye, Werker, & Gerken, 2002). More recently, simulations in-

volving mixture of Gaussians, MOG models, have been used to learn cue weights and fit human behavior fairly accurately (Toscano & McMurray, 2010). These theories concerned how different cues to perception combine to give us the ability to categorize phonemes, but they did not address the degree of categorical effects on perceptual warping. If different cues are more or less prone to categorical effects, and then their individual predictions are weighted to arrive at the category, then the categorical effects would propagate up and the degree of categorical effects in a particular phoneme perception task would be a combination of the effects for the relevant cues. This is quite possible as detection of different cues may be prone to different amounts of interference in the form of noise variance. In fact, Holt and Lotto (2006) found that by varying the variance for a reliable cue in a perception task, listeners would start using a different cue to identify phonemes, suggesting that increased noise variance was causing them to rely on the categorical predictions of another cue, much like the model would suggest [6]. If this account holds up, then there would be repercussions for how L2 learning and perception would occur and the degrees of categorical effects we would expect there, similar to the L2 perception and SLA considerations in the cue weighing for phoneme categorization literature discussed in Holt and Lotto (2006) and Lotto and Holt (2011).

Going beyond these considerations, my goal in this dissertation is not to just use the model to account for speech perception behavior by listeners in their native language. Rather, my goal is to employ this model of speech perception as the ba-

[6]Here we can consider categorical predictions by other cues to be roughly similar to categorical predictions made by knowledge of underlying language phoneme distributions

sis for investigating all perceptual warping in speech perception, whether native or non-native. The basic tenet of this approach is straightforward: the categories we have in our minds affect how we infer intended speech acts by others. As categories develop, whether in infancy or in adulthood, perceptual warping at various points in our phonetic space adjusts, leading eventually to the categorical effects we measure in our behavioral studies. I see two avenues of research that would allow us to see how categories change and to examine the associated differences in perceptual warping. One is to study how infants go from their initial state to adult-like perception. This is outside the scope of this research. The other is to look at adult speakers of one language and to examine changes in their perception as they learn a foreign language. If this model is an accurate description of the relationship between underlying categories, their meaningful, and their associated noise variances and the perceptual warping we see in speech perception, then we should be able to track this relationship as new categories in L2 learning come online. This is exactly what I attempt to do in the next chapter of this dissertation.

Chapter 4: Model L2 Application

In this chapter I examine the use of the Bayesian model described in Chapter 3 for analyzing categorical effects in second language perception. I do this by considering two main situations. First, I use the same type of one-dimensional simulations that I used in the previous chapter to examine language continua involving two categories from different languages. These simulations show that the model is able to capture categorical effects for categories from different languages and also provide insight into how the development of an L2 category changes its effects on surrounding stimuli. Second, I extend the model to multiple dimensions in order to capture categorical effects on perception in a more complex space involving multiple native language and foreign categories to see how they interact as people learn a new language. With these simulations I am able to assess the extent to which the current model can capture more complex language stimuli.

First, I provide a brief review of categorical effects in L2 perception and discuss models that have been used to explain the role of native language categories in L2 perceptual deficiencies. Then I highlight some shortcomings in these models and discuss how my Bayesian model can overcome them. I then discuss the L2 simulations in some more detail providing an analysis of which parameters in the

model are most relevant to L2 category development. I proceed to describe a set of experiments I conducted along with colleagues examining identification, discrimination, and goodness judgments of a multidimensional continuum of vowel stimuli by native English L2 learners of Russian. Using the behavioral results of these experiments I then conduct simulations fitting the model to several one-dimensional continua within the experimental data. Finally, I conduct a multidimensional fit of the model to behavioral data and assess the extent to which the current model is capable of providing a meaningful treatment of such data.

4.1 L2 Categorical Effects Missing Pieces

It is a well established fact that perception of sounds changes with experience. This is trivially true of L1 speakers who learn their first language after enough experience with it, but also true of second language (L2) learning. Even though L2 learners of a new language do not achieve native-like fluency if they start after a certain critical period (Lenneberg, 1967), they do get very proficient with a new language and learn to distinguish certain sound contrasts that did not exist in their L1 (J. E. Flege, 1995). This is modulated by closeness to L1 sounds, allophony in L1 (it has been noted that splitting allophones in an L1 into separate categories in an L2 is a particularly hard problem, perhaps the hardest in L2 acquisition (Barrios, 2013)), age of onset, and overall experience. But we do know that these perceptions change with time. I consider this change to be correlated with changes in the representations of the sound categories. Most of the quantitative experiments

in this area have been production studies where the ability of a foreign speaker to produce appropriate values of new contrasts was measured and evaluated for degree of learning and parameters of the production in relation to the L1 and L2 values of close phonemes/phones (J. E. Flege, 1987). While these studies are immensely useful, they do not tell us anything about underlying categories since production changes could be the result of learned motor behavior and not reflect a change in perception abilities. Other studies consider perception of L2 sounds by naive listeners and consider what happens with learning (Best, 1994; J. E. Flege, 1995). Most of this work is related to one of two models of L2 perception: The Perceptual Assimilation Model (PAM) (Best, 1994) and the Speech Learning Model (SLM) (J. E. Flege, 1995). An overview of these models was provided in Chapter 2, but I will revisit them briefly in order to motivate the experiments and simulations in this chapter.

The PAM predicts how a naive listener of a second language will be able to discriminate contrasts based on how these phonemes map onto phonetic categories in their L1 (Best, 1994). Particularly, the model makes predictions for degree of accuracy in discrimination based on the identification pattern of a foreign contrast. There are four possible identification patterns relating L2 contrasts to L1 categories: Unassimilated (U), 2-category assimilations (2C), category-goodness assimilations (CG), and single-category assimilations (1C). These, in order, lead from the best to the worst discrimination ability. It is worth noting that while these predictions are based on mapping L2 to L1 categories, no explicit metric of measuring phonetic distance is established.

The SLM attempts to account for the ability, the rate, and eventual level of attainment of non-native contrasts (J. E. Flege, 1995, inter alia). Particularly, the learnability of a contrast is determined by its mapping to L1 categories and based on the assumption the an L2 learner uses the L1 as a template (or filter) and then applies similar learning techniques as does a child. According to the SLM, a learner is capable of learning new contrasts, but this ability is modulated by the shared phonological space with the L1, depending on both the time and quality of input. The core predictor of learnability is the dissimilarity of an L2 sound from the closest L1 counterpart, with greater distance leading to greater acquirability. When a new category is not learned, the L2 sound is assimilated to the closest L1 sound. While there is this idea of a "closest" category, like the PAM, the SLM does not make any specific claims as to the nature of the perceptual distance between L1 and L2 sounds. While the SLM does appeal to category change and assimilation, it makes no concrete predictions about the rate of category formation or particular properties of the category that become more precise over time.

There are several shortcomings that these models share when it comes to examining categorical effects on perception. One is that these models do not formally specify perceptual distance or any similarity measure which can be used to make predictions. Another is that these models do not provide for a gradient representation of effects. Then there is the fact that these models only consider endpoint stimuli in a foreign contrast and not usually a full continuum, let alone a continuum between an L1 and an L2 category. Finally there is the lack of an ability to make concrete predictions about behavior based on the language of the speaker without

a priori knowledge of the speakers' identification performance.

By using this Bayesian model to investigate L2 effects in speech perception, I can quantify the perception of the non-native L2 contrasts, thereby validating the qualitative predictions of the PAM and SLM, and providing a detailed view of the particular perceptual measurements and categorical change over time. Further, we can use the formal model to investigate particular effects of relevant parameters, such as category center, variance, and prior probability of category. But, perhaps more importantly, using the same model for analyzing L1 and L2 categorical effects reinforces the idea that these effects are all the result of a common speech recognition procedure, with no need to appeal to any particular cue representation, processing mode, or language context to explain the range of results found by previous studies.

4.2 Application of the Model to L2 Data

The second goal stated in the previous chapter was to employ this model of speech perception as the basis for investigating all perceptual warping in speech perception, whether native or non-native. Instead of appealing to L1 or L2 models, we want to treat perceptual warping in any situation as the result of underlying categories acting on the input. Therefore, as categories develop, whether in infancy or in adulthood, perceptual warping at various points in our phonetic space adjusts, leading eventually to the categorical effects we measure in our behavioral studies. Here I will consider how this model can be used directly to examine L2 effects.

In native language simulations in chapter 3, the listener was modeled as infer-

ring target productions of the speaker using the same set of categories. In the case of cross-linguistic perception, the underlying categories of the listener are not the same as the speaker. In fact, we do not really need to involve a "speaker" at all, and instead focus on the perception of any arbitrary sounds by a listener of a particular native language. Then if we consider L2 learners of another language, it is just a matter of finding what additional categories account for the observed behavior. The case of discriminating non-native contrasts then amounts to just consider the degree of discriminability in the area of the sounds being considered, or comparing the perceptual distance between these sounds to other sounds in the perceptual space. Perception can be investigated by giving the listener a range of stimuli in a given phonetic space and having them complete an identification and discrimination task. When different groups of language speakers complete the same tasks for the same range of stimuli, we can examine category parameter changes between groups by fitting the model to the behavioral data. This provides a measure of how the categories change over time if we consider monolingual speakers as well as L2 learners throughout the learning process.

We can investigate perceptual warping and underlying categories, but we can also focus on building upon findings using the PAM and SLM. We can do this by investigating discrimination of sounds that would be predicted to be assimilated into either single categories, multiple categories, or stay unassimilated. Fitting the model would both validate qualitative findings as well as provide a range of quantitative results that would give us insight into the properties of perceptual distance and discriminability that guide the differences between these different cases.

But critically, given the fluidity of the model, we would get a true sense of the range of effects without having to artificially bin the different scenarios into 3 or more categories of relationships between L1 and L2 contrasts, while still allowing for phonology to play a role along-side fine acoustic measures.

As a first step, applying the model to monolingual data provides an initial snapshot of the parameters that the model derives, particularly the category means, the meaningful category variance, and the noise variance associated with a particular category. Then, the model is fit to behavioral data over the same stimuli on people at different stages of L2 learning. By considering the same parameters, we can see exactly what is changing over time and what is "sub-optimal" about the L2 category representations. Ideally, this fitting procedure would show how the category representations approach those of the native speaker over time while examining precisely which aspects of the category representation are changing. Within the scope of our model there are multiple possible sources of sub-optimal performance when it comes to the perception of non-native contrasts. Each of the four parameters - category mean, underlying category variance, perceptual noise variance, and prior category probability - can vary in different ways, and interact with each other, to create non-native behavior. It is precisely this ability to pull apart these different parameters and consider the implication for a theory of category change with learning that makes the qualitative and quantitative fits and predictions of this model an important contribution beyond past work. In the next section, once I describe the stimuli and the language choice for the experiment, I will in turn consider precisely how each parameter may turn out.

4.3 Experiments 1-3: L2 Learner Identification, Discrimination, and Goodness

In this section I describe a series of experiments designed to collect the necessary L2 learner data as well as both L1 and L2 native speaker control data that is needed in order to use the model to investigate categorical effects in L2 perception. An identification, a discrimination, and a goodness judgment study were completed examining vowel perception in the two-dimensional continuum with endpoints at /i/, /ɪ/, and /ɨ/. The groups were L2 learners of Russian as well as native English and native Russian control groups. Below I describe language selection, stimuli creation, possible effects of individual model parameters, experimental procedures, and an overview of results. Actual model fits and parameter extraction is covered in the next section that covers all model simulations.

4.3.1 Languages

The language for this experiment needs to have at least some phones or phonemes that do not appear in English, but can be measured along a continuum with English phones. Because we want to examine multiple categories in multiple dimensions, it makes sense to consider vowels along the F1/F2 space. What we would like is a language with a vowel in a section of the vowel chart that does not appear in standard American English and is not nestled between other vowels in the same language, so as to eliminate their effects when considering categorical ef-

fects. Russian fits this profile quite well. In Russian there is the /ɨ/ vowel, a high central unrounded vowel, that is in a space of the vowel chart unoccupied by any English vowel. This means that learners of Russian would have to learn a novel category, and we do not know a-priori what their classification of the vowel will look like in relation to existing English categories. Further, Russian lacks the /ɪ/ vowel, which means that the experiment can later be run in the other direction, examining Russian-speaking learners of English and how they learn the /ɪ/ vowel. However, this would have a confound of the /ɪ/ vowel being rather close in terms of first and second formant values to the English /e/ vowel.

Beyond the particular vowel qualities that make Russian a good language for the experiments, there is also the factor of the language-speaking population. First, Russian-speaking learners of English as well as English-speaking learners of Russian are both fairly accessible groups. For me they are highly accessible as I am a native speaker of English. Second, for English-speaking learners of Russian, there is a very high probability that they have had no to minimal exposure to the Russian language, or even to the Slavic family of languages, prior to formally studying the language. This is not necessarily true in the other direction, as most Russian speakers have had a fair amount of exposure to English. However, this would be a concern for most foreign languages with accessible populations.

There are some caveats to the vowel relationships that I cite above worth mentioning. I stated that Russian speakers do not have a category for the English /ɪ/ and that English speakers do not have a category for the Russian /ɨ/. However, this is not as clear cut as it may seem. First, there is reason to believe the English

speakers have something similar to the Russian /ɨ/ as a variant (or allophone) of a schwa, /ə/. This is heard as the first vowel in the word "in" in the phrase "just in time". While this is based on a particular analysis of the phone and also would represent a rare form thereof, it is still worth bearing in mind as a potential source of better than expected performance by monolingual English speakers. For Russians, there is a good chance that they will have an exposure to the English /ɪ/ if they are themselves speakers of, or have a strong familiarity with, Ukrainian. In Ukrainian, the /ɨ/ vowel manifests with spectral values much closer to the English /ɪ/. Hence, we will need to be careful with what familiarity with other Slavic languages our Russian speakers have. Finally, there is a dispute as to whether /i/ and /ɨ/ are separate phonemes or allophones of the same phoneme. The Moscow school of linguistics believes that they are allophones, citing their almost complimentary distribution. Meanwhile, the St. Petersburg school of linguistics maintains that they are separate phonemes. It is possible that this has implications for the behavior of the speakers in the experiment. However, since the phones are so perceptually distinct, and have at least one minimal pair, and are taught in schools as separate vowels, I do not foresee this as being a major issue.

4.3.2 Stimuli

The stimuli for the experiment constitute a two-dimensional continuum of vowels covering a triangular area encompassing the vowel space around the vowels /i/, /ɨ/, and /ɪ/. The end-points of the triangle are all judged to be good exemplars

of these three vowels. The range of F1/F2 values was chosen based on previously published values as well as verification via recording and analysis of a native English, native Russian, and bilingual English-Russian speakers. Below are representative values for these vowels from the literature.

English Vowels					
	Male		Female		
	F1	F2	F1	F2	Source
/i/-English	342	2322	437	2761	J. Hillenbrand et al. (1995)
/ɪ/-English	427	2034	483	2365	J. Hillenbrand et al. (1995)

Russian Vowels			
	F1	F2	Source
/i/-Russian	150	2400	M. Halle and Jones (1959)
/ɨ/-Russian	200	1600	M. Halle and Jones (1959)

In order to largely cover the needed area I varied F1 from 200-450 and F2 from 1550 to 2350. While this did not cover the most extreme values for the vowels, it captured enough of the space to clearly range between good exemplars of all categories. Equal steps between stimuli were measured in Bark steps, a psychologically plausible log-based acoustic measure. In Barks, the continuum ranged from 2.5 to 4.5 for F1 and from 11.5 to 14 for F2. In Table 4.1 stimuli F1/F2 values chosen for the experiment are marked by the stimulus id based on the row and column in the continuum (for handy reference later in this work).

Pilot testing showed that people were more accurate at identifying the vowels if they were not standalone but rather in a VC syllable, so a voiceless /p/ burst

Barks (Hertz)	14 (2319)	13.69 (2213)	13.38 (2112)	13.06 (2016)	12.75 (1924)	12.44 (1836)	12.13 (1752)	11.81 (1672)	11.5 (1595)
2.5 (250)	1_1		1_3		1_5		1_7		1_9
3.0 (297)		2_2		2_4		2_6		2_8	
3.5 (347)			3_3		3_5		3_7		
4.0 (399)				4_4		4_6			
4.5 (453)					5_5				

Table 4.1: F1 and F2 values for experimental stimuli for Experiments 4-6. Values across the top are F2, values on the left vertical are F1. Values are provided in Barks, with corresponding Hertz values in parentheses

was added to each vowel for the experiment. Initial vowels were synthesized with steady formant structures based on midpoints in human productions. However, pilot testing showed that people had trouble doing proper identification, possibly due to the presence of diphthongization in both the /ɪ/ and /ɨ/ in natural productions. Because of this, a novel procedure was created for creating stimuli that had full diphthong structures, while still reflecting all central measures for vowels at all points in the continuum. This procedure is detailed in the following section.

4.3.2.1 Stimulus Creation Process

The two-dimensional vowel continuum was created using a modified version of the LPC decomposition and resynthesis procedure in Praat. In this procedure, a human production is inverse filtered to create a speech-like sound without any vowel quality in the form of formant resonances. Then, this sound is re-filtered with specified formant grids, creating a sound with any desired formant structure while

maintaining the original production's voice quality and temporal properties. Stimuli in this experiment were based on an initial /ɪ/ production and the LPC analysis was carried out using the procedure described in (Winn & Litovsky, Submitted).

The formant grids for the re-filtered experimental stimuli were created via linear combination of three ideal formant structures extracted from human productions of /i/, /ɪ/, and /ɨ/. These productions were made by an adult bilingual Russian-English speaker. For each stimulus in the 2-dimensional experimental grid, weights were calculated for weighting the three ideal sound tokens by computing distances between the stimulus and the three tokens. First, we find a line through the stimulus and one of the three ideal vertices and see where it intersects the opposite side of the ideal token triangle. We find the relative contribution of the other two ideal tokens to this intersecting point and then scale based on how far the stimulus is along the line connecting this intersection and the third vertex. Hence we get the weights that we use to combine formant grids for the three ideal tokens. See Appendix F for a full listing of weights and formant grids.

These weights are then applied across the entire duration of the vowel to create the formant grids for the experiments. These formant grids are then used to re-filter the residual sound source from the LPC decomposition to create the actual stimuli used in the experiment. The experimental stimuli were designed to vary in the first and second formants, so a two-stage process was implemented in order to combine the lower energy structure of the re-filtered stimuli together with the higher order energy from the original recording. This also ensured that as much of the original sound remained in all the stimuli, minimizing the differences between them. A

low-pass 2300 Hz Hann filter was applied to the re-filtered stimuli with a buffer that linearly shrank to 0 energy between 2300 and 2500 Hz. A high-pass filter was applied to the original /ɪ/ recording discarding all energies between 0 and 2300 and linearly growing to full energy by 2500 Hz. Because timing information was based on the original recording, these two components come together naturally, producing high-quality stimuli that contained manipulated F1 and F2 while keeping all other formants intact. Once vowels were completed, the closure and release of a single /p/ consonant was spiced onto the end of each re-filtered vowel to maintain uniformity across the stimuli set. Subsequent pilot tests showed that both native Russian and native English speakers reliably identified all endpoint stimuli based only on the varying F1 and F2 structures.

4.3.3 Potential Influence of Different Parameters

The overall goal of this chapter is to gain a better understanding of changes that happen in speech perception over the course of L2 acquisition through the lens of the Bayesian model. Hence, there are a specific set of parametric changes that we might be able to see when we look through this lens. In other words, there are multiple ways of capturing the sub-optimal performance of early language learners and the changes that happen with learning. Each of the four parameters in the model — category mean, underlying category variance, perceptual noise variance, and prior category probability — can vary, interact with other parameters, and change over time, leading to various sorts of non-native behavior. Beyond just considering how

such parameters might affect the model fit, its important to acknowledge that these parameters really are expected to be different for naive English speakers and the L2 learners of Russian. Particularly, while in the L1 simulations we were able to assume equal priors, this might not be as accurate an assumption for speakers with changing categories, as learners exposed to a new category might just not be expecting to encounter this category as often. However, I don't believe this is a critical issue in an experimental setting since people are explicitly told they will be hearing tokens from all categories. The fact that the non-native learners will have different means and variances almost goes without saying, as the discussion on L2 effects above illustrated. However, beyond just the settings for these parameters, there is the additional question of robustness of the representations of the perceptual features/cues used to identify the particular phonemes. This idea of robustness is beyond the scope of the present work, and something I consider for future investigation. Meanwhile, in this section I go through each parameter to examine the ways in which it might be playing a role and suggesting what a proper analysis might be to investigate the parameter's contribution.

4.3.3.1 Category Center

For native monolingual English speakers, I do not expect them to have any meaningful mean for the Russian /ɨ/ category. However, it is possible that as a starting point, at least for experimental purposes, they would treat all tokens that are bad exemplars of either /i/ or /ɪ/ as belonging to this third 'Other' category.

In this case, over the course of the experiment they might treat the center of these other tokens as a center of the third category, thereby mimicking appropriate behavior. More importantly, we can consider what could happen to the mean of the third category once English speakers actually learn more Russian. If English speakers assimilate tokens of /i̞/ into either the /i/ or /ɪ/ English category, then they might dissimilate the new category and overshoot the true value, leading to a category centered farther away than it should be. As English speakers gain Russian proficiency, and are exposed to more input in Russian, I expect the center of the category to move toward the actual category center, closer to the representation that native Russian speakers have. When it comes to centers for the two native English categories, I would not expect these to change significantly as English speakers learn Rusian. While there is some research in SLA that has shown native categories changing to become closer to L2 categories over time, these scenarios typically relate to highly proficient L2 speakers, beyond the level of proficiency I am testing in my experiments.

4.3.3.2 Category Variance

For the English speakers, we would not expect the category variance for /i/ and /ɪ/ to vary greatly with Russian proficiency. And we can use the variance of the Russian speakers' /i̞/ category as a baseline to see how that of the English speakers compares. For native English speakers with no Russian we might expect the variance that the model extracts to be very large if they are treating this as an

'Other' category. However, it is possible that given the design of the experiment that they end up treating this third category as a very specific barely variable stimulus. Either way, I would not expect this to influence their perception since it is not a legitimate speech category. For learners of Russian, there are two patterns that are possible when it comes to the category variance. One is that due to a lack of exposure to Russian, learners are not sure about what constitutes good tokens of the vowel category and therefore treat it as a category with huge variance to cover a large amount of vowel space. In the model, this large variance would lead to very little warping of the perceptual space and rather lead the listeners to perceive things quite accurately according to the sounds' acoustics. In terms of identification, it would lead the listeners to identify a larger percentage of stimuli far away from /ɨ/ as this third category since the category would cover so much of the vowel space. Over time, these learners would then reduce the variance of the category once they were exposed to more and more Russian input and got a better idea of the appropriate spread of possible legitimate /ɨ/ values. The second alternative is that initially the English speakers would form a very low-variance category for the Russian /ɨ/ because of the lack of variance in the tokens and contexts to which they have been exposed. This would lead to strong perceptual warping, but only close to the center of the category, since due to the low variance, the probability of categorizing a sound into this category would quickly go to zero as we move away from the center. Over time, these learners would increase the category variance to eventually cover the appropriate space in the vowel space as they are exposed to more and more instances. Extracting category parameters from the identification

and discrimination scores for speakers of varying proficiencies will inform us as to which pattern underlying changes in L2 perception.

4.3.3.3 Perceptual Noise

Predictions with perceptual noise are less straightforward than category means and variances. It is possible that as learners start on a new language, they make an assumption about the center of a category and then treat most of the variability as perceptual noise, not trusting themselves to pull meaningful information from the variability. In this case, they would be very biased toward category centers, and we would expect quick and strong perceptual warping around the new category. Alternatively, the learner might be paying particular attention to the minor acoustic differences in the new vowels as they try to learn a proper representation for the new phoneme. In this case they would treat the task of speech perception as an inherently low-noise problem, and hang on to acoustic details. This would lead to smoother discrimination functions with little if any peakedness along the continuum to indicate warping. There is another possibility, especially with early learners. Because many of the stimuli played for the participants will be between endpoint vowels, they may treat the whole task as a noisy problem, artificially increasing pulling toward category centers. This may particularly increase pulling toward the English category centers since they would be the most robust ones for the listener. On the up side, this would cause more perceptual warping at the center of the continuum allowing us to finely investigate how the categories interact with each

other.

4.3.3.4 Prior Probability

The prior probability of the categories may play a smaller role in an experimental setting than a real life setting. Because the stimuli are presented in isolation and there is no co-articulation or larger context, the fact that one category is more likely in real life may not carry over into the experiment. In other words, participants know they are hearing sounds from three categories, and therefore might be resetting their expectations for just the experiment. If this is the case, then we can keep the equal prior probability assumption in place from the previous chapter's simulations. On the other hand, the learners know that this experiment is being conducted because they are studying Russian. This may cause them to ramp up their expectation for the Russian vowel category. This would in turn lead to a higher classification of items into /ɨ/ and a shift of the phonetic boundary toward /i/ and /ɪ/. It is possible however that the participants will have a higher prior for their native English categories since they are more accustomed to hearing them. This could lead them to hypothesize a category center of /ɨ/ closer to the English vowels in order to still classify some things as /ɨ/s. No matter the prediction, there is a tight interplay between the prior probability of a category and the variances of that category. During analysis we will have to be careful to pull their separate effects apart where necessary.

4.3.4 Tasks and Procedure

Participants in the experiments conducted an identification, a discrimination, and a goodness rating experiment. All participants started with the discrimination task, then completed the identification task, and finished with the goodness tasks. There are training concerns no matter what order the tasks are done in, and sometimes counterbalancing is the way to go, but we wanted to ensure that discrimination performance was in no way colored by the experience from the other two tasks and so kept the task order constant. Additionally, participants filled out a language background questionnaire and recordings of a Russian paragraph of text were collected from each participant for future analysis. The questionnaire was used in order to assign people to the proper language category for analysis purposes. The recorded paragraph was designed to elicit the maximum number of Russian phonemes and consonant clusters in order to later collect accent judgments by native Russian speakers. In the future this will be used to correlate production and perception of non-native vowel categories, but lies outside the scope of the present work. The identification, discrimination, and goodness judgments are the ones we are primarily concerned with here.

4.3.4.1 Participants

There were four experimental groups and a control group for this set of experiments. The experimental groups included native English speakers, and beginner, intermediate, and advanced learners of Russian. The control group consisted of a

group of Russian speakers comprised of both native and heritage speakers[1]. There were 16 native English speakers (10 female, mean age = 20.3), 16 beginner learners of Russian (9 female, mean age = 20.9), 9 intermediate learners of Russian (6 female, mean age = 25.8), 19 advanced learners of Russian (8 female, mean age 28.3), and 7 native speakers of Russian (5 female, mean age 21.6). Participants were recruited from both the University of Maryland and from the University of Utah. The breakdown of participants from the two locations is presented in Table 4.2. Data from a total of 13 participants (4 Native English, 4 Beginner, 1 Intermediate, 2 Advanced, and 2 Russian Native) was excluded from the final analysis because the participants mixed up the mapping between categories and labels in the identification portion of the experiment[2].

The participants were assigned to language groups based on responses to the language background questionnaire. Native English speakers were identified as those who spoke dominant English from birth and have had no formal education in Russian, have had no exposure to Russian through travel or relatives, and have not formally studied any other Slavic languages close to Russian. The Russian group was comprised of those who spoke Russian dominantly at home from birth and either remained living in Russia or moved to the United States after puberty but continued

[1]All heritage and native speakers self-reported high proficiency across the board and had no trouble with pronunciation of any Russian speech on the production portion of the experiment. While there are certainly differences in proficiency between the native and heritage speakers, this contrast is beyond the scope of the current work. More importantly, both serve as a good baseline of having a fully formed category for the Russian vowel /ɨ/

[2]This is a high number of participants to exclude. This is due to the identification task design, where the categories were labeled 'Category 1', 'Category 2', and 'Category 3' as opposed to the names of the actual vowels. While the task allowed us to keep the directions constant across all participant groups, it did lead to an unexpectedly high level of confusion. Overall this constitutes a 19% exclusion rate

	UMD - College Park (CP)		University of Utah (Utah)	
Level of Russian	Female	Male	Female	Male
Naive English	10	6	0	0
Beginner	2	3	7	4
Intermediate	2	1	4	2
Advanced	5	3	3	8
Native Russian	5	2	0	0

Table 4.2: Distribution of participants for Experiments 4-6

to use Russian actively in the home. The Russian learners were split up into Beginner, Intermediate, and Advanced groups based on a combination of years of formal education in Russian and whether they had been immersed in a Russian-speaking environment for more than a year. Beginners were first year students of Russian who had no immersion experience. Intermediate learners were in their second year of learning and have had no immersion experience. Advanced learners were those with more than 2 years of formal Russian instruction or people who had only 1 or 2 years of formal instructions but had been in a Russian immersion experience. Most of these came from the testing center in Utah where they were formally studying Russian upon the completion of their Church of Latter Day Saints mission in Russia.

All participants in the experiment gave informed consent and received either course credit or a $10 payment for their participation. The experimental session lasted approximately 1 hour.

4.3.4.2 Technology

Experiments at both locations were conducted on a MacBook Pro 2009 laptop running Psyscope experimental suite version B70 (http://psy.ck.sissa.it/) [3]. The particular recording equipment used at the two sites varied slightly. In Utah, participants were recorded inside a WhisperRoom Model 4242E sound insulation booth. They listened to stimuli bilaterally over Sony MDR-7506 Professional model headphones. Button responses were recorded via external keyboard and audio recordings were captured with a Samson QV10E headset condenser microphone and recorded onto a Marantz PMD661 Handheld SD Recorder. In College Park, participants were tested inside a regular sound-insulated experimental room at the Linguistics Department. Participants listened to stimuli over Sony MDR-V7 dynamic stereo headphones and button responses were recorded directly on the laptop keyboard in front of them. Participants' spoken responses were captured with a RODE NT1 XLR microphone via XLR Blue Icicle preamplifier at 16-bit/44.1 kHz resolution and recorded directly onto the Praat Speech Analysis Software version 5.3.69 (Boersma, 2001). The digital stimuli for playback were sampled at 44.1 kHz with a 16 bit quantization rate and mean power intensity of 75 dB.

[3]This version of Psyscope contains a bug with the randomization of the stimulus presentation. While the actual stimuli in the experiment are presented in a truly randomized order for any given list, the log files always report that the first element in the original declaration file is the one that was presented first, even though it was actually another randomly selected element. Because of this, there was a small additional post-processing step necessary to calculate the item that was presented once less than all others in order to reconstitute the proper log files for analysis. This bug persists to the current version of Psyscope to the best of my knowledge as of October, 2014.

4.3.4.3 Identification

The identification task was a 3-way forced choice identification task. For every stimulus heard, the participants had to classify it into one of three categories. The categories correspond to the English categories /i/ and /ɪ/ and the Russian category /ɨ/. However, native Russian speakers are not familiar with the English category /ɪ/ and native English speakers are not familiar with the Russian category /ɨ/. To get around this problem, and to make sure that all participants completed the same experiment, I simply refer to these categories numerically as "Category 1", "Category 2", and "Category 3". The original idea was to have everybody complete the same experiment so that we could use the native performance as a baseline to see how it changes over time. However, it turned out that having English speakers classify sounds into a Russian category was not very informative since they treated the task differently from the Russian learner. In the end, it would have been more valuable to have native English speaker assimilation data just into the native English categories to see how they were assimilating the Russian /ɨ/. This is definitely a follow-up experiment that warrants running in the future.

After receiving experimental instructions on the screen, participants went through a training/learning phase where they associated sound exemplars with the three category labels before proceeding. The participants got to listen to an example from each of the three categories up to 5 times by pressing the buttons 'd', 'g', or 'j' for categories 1, 2 or 3, respectively. The exemplars used for these categories were the corner stimuli that received the highest goodness ratings as members of

these categories, stimuli 1_1 for /i/, 1_9 for /ɨ/, and 5_5 for /ɪ/. One concern here that I will address later is that for the native English speakers who lack a Russian category /ɨ/, they may have trained on the exemplar and used a purely acoustic classification metric during the experiment, leading to data that does not lend itself to easy analysis. This is not unreasonable, since listeners have been shown to be able to employ both an acoustic and phonetic mode of processing in past experiments (Pisoni, 1973).

Once the learning phase was complete, participants went on to the main experiment. In this part of the experiment participants heared each of the 15 stimuli 10 times in a randomized order. Each trial started with 500 ms of silence and then the stimulus was played. The trial continued until the user pressed the 'd', 'g', or 'f' button, with a 3000 ms timeout. The inter-stimulus interval was randomly sampled from 250, 500, 750, and 1500 ms.

4.3.4.4 Discrimination

The discrimination task was a classic AX forced choice experiment. For each pair of stimuli, the listeners had to select whether the two speech sounds they heard are the same or different. Because of the number of stimuli in the 2-d continuum, we could not test all combinations of pairs. Hence, I only used 1 and 2-step pairs throughout the whole continuum. However, 2-step does not necessarily mean following two stimuli in the same direction. For example, consulting Table 4.1, stimuli 3_5 and 2_8 are separated by 2 steps, one to the upper right and then one horizontally

over. All together there were 66 different pairs of stimuli that I included in the experiment.

The experiment started with instructions for the participant and then a practice section where the participants could get used to hitting the keys for same or different pairs. The practice section consisted of 10 trials, with some SAME, a few immediate neighbors, and some very distant pairs to ensure that some were perceived as being different. Each practice trial started with a 150 ms delay, contained a 200 ms ISI, and used a 500 ms ITI.

Then the participant conducted the actual experiment. The experiment consisted of 177 trials, repeated twice in random order. The 177 trials were comprised of 66 different pairs 2 times each (in AB and BA order), and 45 same trials (15 stimuli, 3 times each). Each trial started with a 150 ms initial delay and had a 200 ms ISI. To keep the timing variable, the ITI was sampled randomly from 565, 624, 783, and 943 ms. Upon hearing the pair of stimuli, the participants had up to 6000 ms to respond with one of the two buttons, 'f' for same or 'g' for different. Before analysis, trials that fell on either extreme of the reaction time distribution were thrown out. Out of 17,685 total trials, 66 with reaction times below 500 ms and 193 with reaction times over 3000 ms were removed (1.46% loss rate).

4.3.4.5 Goodness

The goodness task employed a simple 5-point Likert scale to gauge goodness of fit of the stimuli to the three categories in the experiment. This experiment

was split up into three sections, one for each of the three vowels. In each section, the participant received instructions on the use of the Likert scale and then heard three exemplars of good members of this category. The good exemplars were always the endpoint stimuli from the triangle. For English speakers learning Russian these served merely to remind them which of the three categories they are judging on. For native English speakers and native Russian speakers, this also gave them a reference point for the third vowel category that does not exist in their language.

After the familiarization phase, participants went on to the main part of the experiment. They judged each stimulus in the 2-d continuum two times. Each trial started with 50 ms of silence, then the item was played and the participant was given a chance to respond by pressing 1, 2, 3, 4, or 5 on the keyboard with a 3000 second timeout. Once a response was received there was 250 ms of silence followed by a beep and the experiment would advance to the next trial. This was repeated for each of the three vowels, ending up with 90 total judgments of category fit (3 vowels $\times$ 15 stimuli $\times$ 2 repetitions).

4.3.5 Results

In this section I present the results of the experiments without appealing to the parameters extracted by the model.

4.3.5.1 Goodness Results

The goodness experiment confirmed that the categories were being perceived the way I expected based on pilot testing. The primary point of interest is that the corner stimuli at the three endpoints of the 2-dimensional continuum were accepted as very good exemplars of the three categories by the relevant native speakers. Also, to make sure that people were not just rating everything as good stimuli no matter what vowel they were supposed to represent, its important to check that the corner stimuli are also judged as bad examples of both the other two vowel categories.

Figure 4.1 shows the distribution of goodness ratings by native English speakers for the three corner stimuli as all three of the possible vowels. Primarily I note that English speakers judged the 1_1 stimulus as /i/ and the 5_5 stimulus as /ɪ/ almost unanimously 5 out of 5. At the same time, the 1_1 and 5_5 stimuli were judged as really bad exemplars of both of the other vowel categories. The only case that is not entirely clear is English speakers judging the top right corner stimulus, 1_9, which is a prototypical /i̟/ as instances of /i/. While we still see a plurality of 1 out of 5 ratings, we do get poor ratings across the board, suggesting that native English speakers have some difficulty judging this stimulus as definitely being separate from the /i/ category. This is an early indication that English speakers may be assimilating the /i̟/ category into their native /i/ category.

Figure 4.2 shows the distribution of goodness ratings by native Russian speakers for the three corner stimuli as all three of the possible vowels. Primarily I note that Russian speakers judged the 1_9 stimulus as /i̟/ unanimously as 5 out of 5.

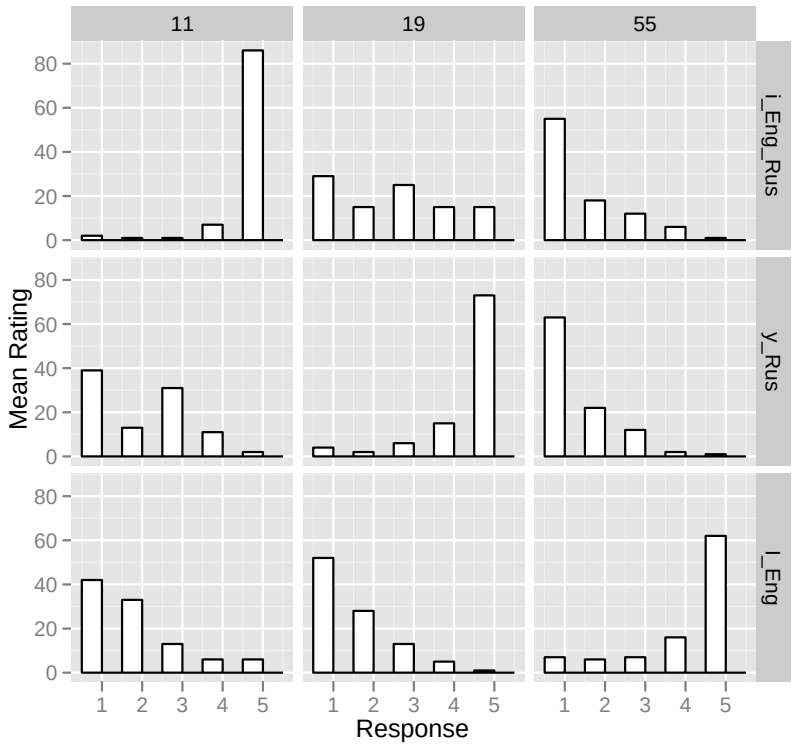


Figure 4.1: Distribution of goodness ratings by naive native English speakers. Judgments are on a 1-5 scale for each of the three corner stimuli as representative of the three vowel categories /i/ (i_Eng_Rus), /ɨ/ (y_Rus), and /ɪ/ (I_Eng). Distributions along the diagonal represent ratings of corner stimuli as the vowel they are meant to represent. The other squares in the plot represent judgments of the corner stimuli as the other two vowels

Further, they judged this corner stimulus as almost exclusively 1 out of 5 for how well it represents either of the other vowels. They had some difficulty assigning appropriate goodness judgments for the English /i/ vowel as a representative of English /ɪ/, which is not surprising given that the /ɪ/ category is non-native for them.

Taken together, the native English and Russian goodness ratings suggest that all native speakers judged their native vowels as being great representatives of the

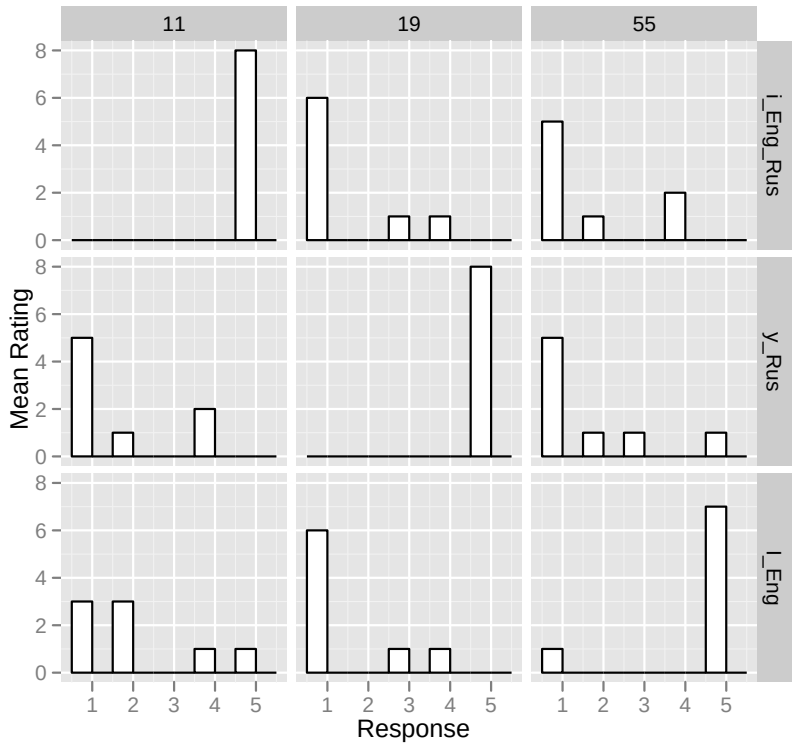


Figure 4.2: Distribution of goodness ratings by native Russian speakers. Judgments are on a 1-5 scale for each of the three corner stimuli as representative of the three vowel categories /i/ (i_Eng_Rus), /ɨ/ (y_Rus), and /ɪ/ (I_Eng). Distributions along the diagonal represent ratings of corner stimuli as the vowel they are meant to represent. The other squares in the plot represent judgments of the corner stimuli as the other two vowels

proper categories. This means that the stimuli were well constructed and also gives some initial insight into how people might be representing these categories, primarily the indication that English speakers may be assimilating the non-native Russian category into their English /i/ category.

4.3.5.2 Identification Results

There are two ways of looking at the identification results from the experiment. One is the overall identification pattern for classification into one of the three categories along the entire 2-dimensional continuum. The second is to consider the three individual continua between pairs of vowels along the outside edges of the stimuli triangle. Figure 4.3 shows the full set of categorization performance for all stimuli in the experiment, across all levels of Participants. We can see that when grouped across all levels, participants in the experiment performed as expected. Along the outside of the triangle, there is a clear transition around the mid-point between the two corner stimuli with very little classification into the third vowel anywhere on the continuum. The central three stimuli exhibit very fast transitions in identification patterns between them, with a different category becoming dominant in each direction from the center of the continuum. This indicates that the center of the continuum is a good approximation of the three-way category boundary in vowel space for our participants. Finally, we can see that while at the /i/ and /ɪ/ corners the classification is almost unanimous, there is a lot of classification as /i/ and almost no classification as /ɪ/ at the /ɨ/ corner of the continuum. Together with the goodness results, this indicates that the English-speaking learners of Russian are assimilating this vowel into the English /i/ category. Based on the predictions made by both the PAM and SLM L2 speech perception models, we would expect that English speakers would have increased difficulty discriminating tokens of /i/ and /ɨ/ but not tokens of /ɪ/ and /ɨ/. In our experiments and simulations, I extend

this idea to predict that when viewing the continua between these pairs of vowels we would get greater d' scores near the /ɨ/ category when considering the continuum from /ɪ/ than when considering the continuum from /i/.

When we consider the probabilities of identification just along any of the borders we can examine performance across different levels of proficiency. Figures 4.4-4.6 show the patterns for identification, along with the best fitting logistic to the identification data, for all learner levels for each of the three 1-dimensional continua. From these graphs we can see that the most consistently sharp identification curves are those for the /i/-/ɪ/ continuum for all native English speakers. This makes sense since the English categories are very robust in adults and should have steady categorical effects. The /i/-/ɨ/ continuum exhibits the shallowest identification curve, and for the beginner and intermediate learners never reaches below 30%. However, by advanced stages of learning we see the curve starting to approximate that of native Russian speakers, which have these two categories natively. There is not as clear a pattern with the /ɪ/-/ɨ/ continuum. The beginner learners do have a very shallow curve that gets stronger categorically with learning, but there is no baseline to compare it to since no group of speakers has both these vowels natively. When we look across the different continua here, we can also see that none of the native English speakers show the same inflection point on the identification curves as the Russian speakers. While the difference on the different continua is more or less pronounced, it seems to be constant across all three. Identification curves are known to shift with prior categories and category centers, so this might be an early indication that the underlying categories may vary between the speaker groups. I will consider

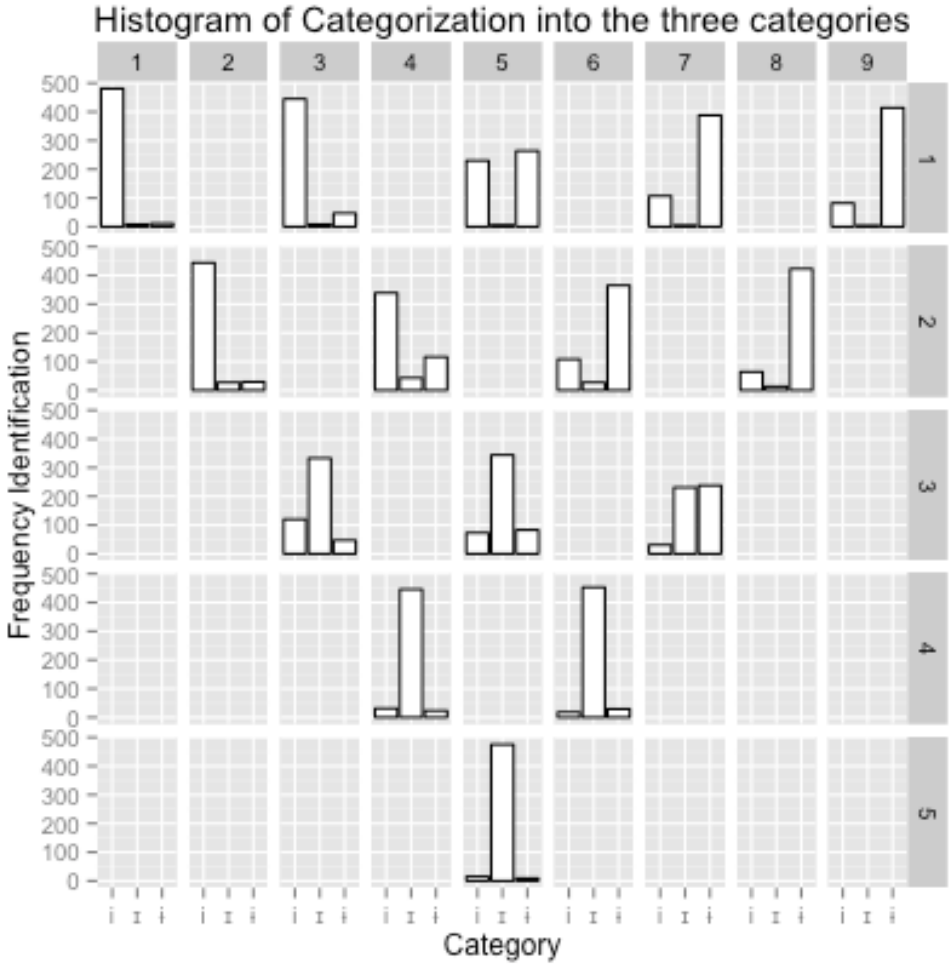


Figure 4.3: Distribution of identification judgments by all participants for all stimuli in the 2D continuum. Each histogram represents, from left to right, classification as /i/, /ɪ/, and /ɨ/.

both the potential differences in priors as well as examine the underlying category means in following sections. In general, this also suggests that even by advanced levels in this experiment, the participants still have a ways to go before they reflect native Russian speaker performance. Full values for all identification results can be found in Appendix G.

4.3.5.3 Discrimination Results

For discrimination, the large number of possible 1 and 2 step pairs in the 2-dimensional continuum lead to a small number of judgments for any individual pair. Because of this, it is impossible to consider d-prime (d') scores for individual participants in analyzing their discrimination performance. Hence, I examine all discrimination results using data aggregated across participants at different Russian proficiency levels. This allows us to get a meaningful d' score to use as a psychoacoustic distance metric for model fitting in the next section. Tables 4.3 and 4.4 presents all 1-step and 2-step d' scores for all levels of Russian learners.

Not all the groups show signs of categorical perception in all the continua. We see this by the lack of a clear peak in discrimination scores in the middle of the continua. Of course, we wouldn't necessary expect strong categorical perception in any case since we are dealing with vowels and many past experiments have failed to show very strong categorical effects for vowels. In fact, it was specifically the lack of a discrimination peak in the middle of the continua, while still having an S-curve type identification peak, that lead some researchers to conclude that vowels do not

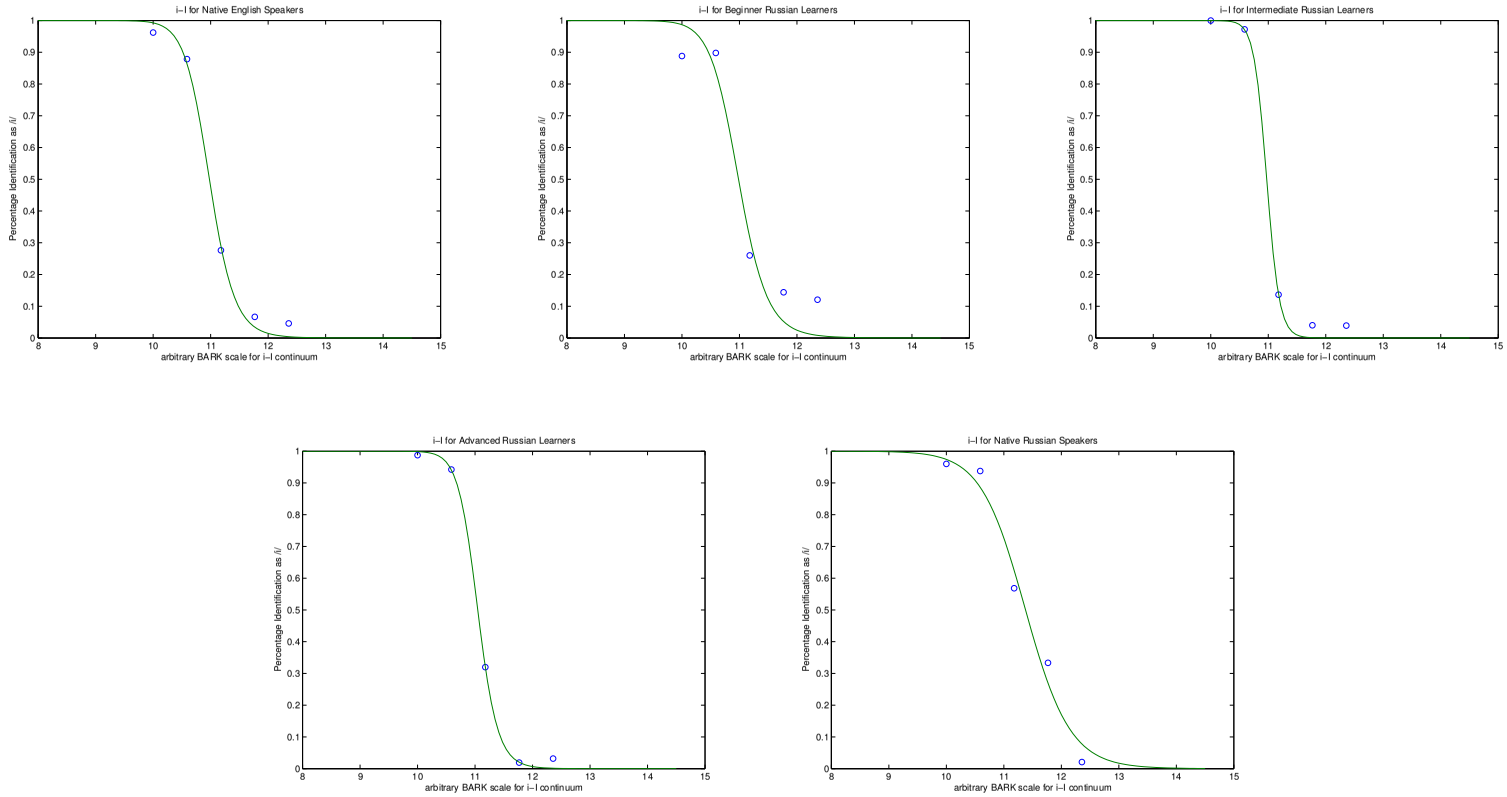


Figure 4.4: Behavioral measures of identification along the /i/-/ɪ/ continuum. Overlaid is the best fit logistic from the proceeding identification simulation section, showing the rate of change of the function and giving an idea of where the category boundary lies. a) Native English speakers, b) Beginner Russian learners, c) Intermediate learners, d) Advanced learners, e) Native Russian speakers

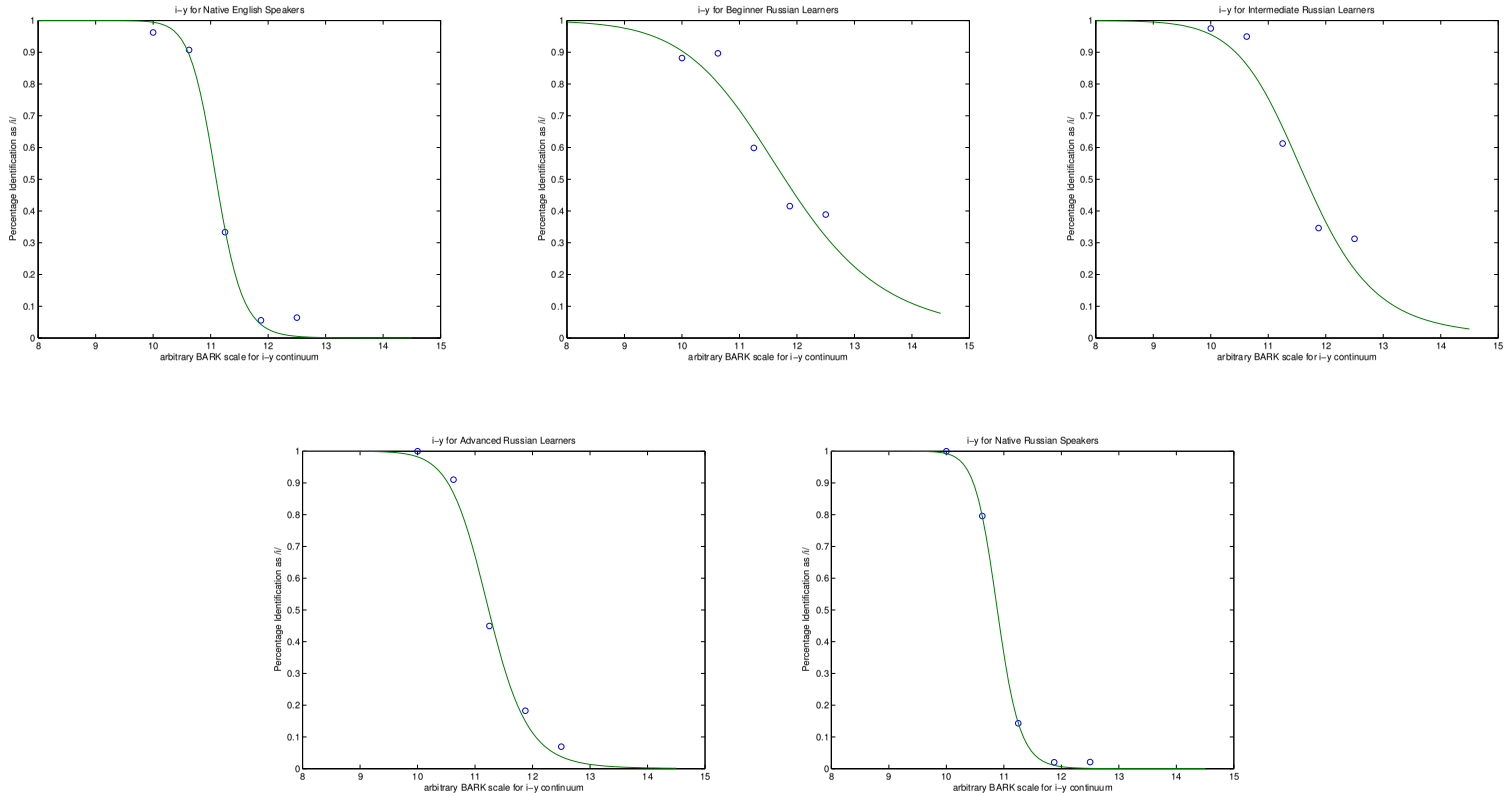


Figure 4.5: Behavioral measures of identification along the /i/-/ɨ/ continuum. Overlaid is the best fit logistic from the proceeding identification simulation section, showing the rate of change of the function and giving an idea of where the category boundary lies. a) Native English speakers, b) Beginner Russian learners, c) Intermediate learners, d) Advanced learners, e) Native Russian speakers

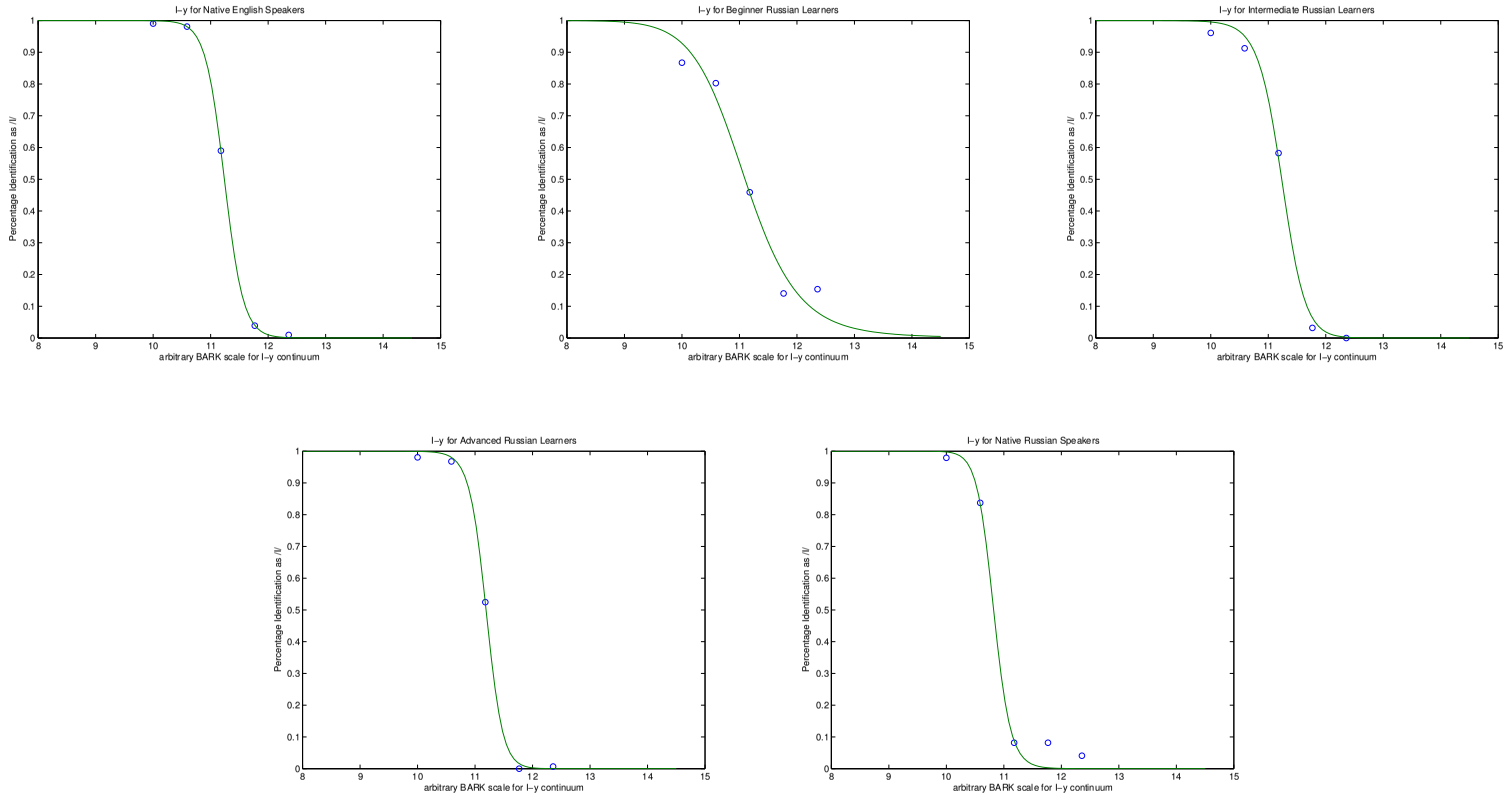


Figure 4.6: Behavioral measures of identification along the /ɪ/-/i/ continuum. Overlaid is the best fit logistic from the proceeding identification simulation section, showing the rate of change of the function and giving an idea of where the category boundary lies. a) Native English speakers, b) Beginner Russian learners, c) Intermediate learners, d) Advanced learners, e) Native Russian speakers

Learner Level	S1-S2	S2-S3	S3-S4	S4-S5
Continuum between /i/ and /ɨ/				
Native English	1.48	0.66	-0.53	-0.14
Beginner	0.67	0.46	-0.08	0.22
Intermediate	0.48	1.44	1.28	0.19
Advanced	1.05	0.94	0.47	0.65
Native Russian	0.88	0.99	0.08	0.1
Continuum between /i/ and /ɪ/				
Native English	1.57	2.51	1.2	1.33
Beginner	0.59	3.08	1.91	0.9
Intermediate	1.44	3.44	0.91	0.39
Advanced	1.33	2.98	1.49	1.11
Native Russian	0.6	0.51	0.28	1.22
Continuum between /ɪ/ and /ɨ/				
Native English	0.89	2.5	1.83	1.82
Beginner	0.49	2.35	1.29	0.72
Intermediate	0.09	2.44	1.33	1.14
Advanced	0.47	2.42	1.47	1.4
Native Russian	-0.16	0.88	0.94	1.18

Table 4.3: 1-step D-prime scores for all levels of learners for the three vowel continua around the sides of the 2-D stimuli continuum

Learner Level	S1-S3	S2-S4	S3-S5
Continuum between /i/ and /ɨ/			
Native English	1.95	2.05	0.88
Beginner	1.78	1.76	1.13
Intermediate	2.39	2.17	1.6
Advanced	2.07	1.99	1.47
Native Russian	2.45	1.91	1.2
Continuum between /i/ and /ɪ/			
Native English	3.68	3.55	2.27
Beginner	4.2	4.32	2.33
Intermediate	3.14	4.16	1.68
Advanced	3.9	4.17	2.66
Native Russian	2.09	1.9	1.08
Continuum between /ɪ/ and /ɨ/			
Native English	2.88	3.92	2.68
Beginner	4	3.88	2.69
Intermediate	4.18	3.94	2.95
Advanced	3.45	3.72	2.42
Native Russian	2.13	2.34	2.21

Table 4.4: 2-step D-prime scores for all levels of learners for the three vowel continua around the sides of the 2-D stimuli continuum

have perceptual warping due to categorical effects (e.g., D. B. Fry et al., 1962).

However, we do see some clear signs of categorical effects. For all groups of English speakers, we have a peak in discrimination at the middle of the /i/-/ɪ/ continuum, as evidenced by both 1-step and 2-step discrimination scores. This is encouraging since this continuum is native in English and categorical effects should be present independently of Russian learning level. Interestingly, native Russian speakers do not show any increase in discrimination along this continuum, displaying poor discrimination throughout. The Russian speakers do show an increase in discrimination toward the /ɪ/ category. This can be explained by the fact that Russian speakers do not have this category and therefore no category center to bias the perception in this area of vowel space.

If we consider the English-speaking groups on the 2-step /ɪ/-/ɨ/ continuum, we see a very strong peak in discrimination in the middle, but very different performance at the two sides. The English speakers are much better at discriminating neighboring vowels near the Russian /ɨ/ category than the English /ɪ/ category. Again, this may be due to the fact that there is no native English category that warps the perceptual space at this end of the continuum.

When we turn to the /i/-/ɨ/ continuum, the d‘ scores we see are pathological, in that they do not correspond with any expectations. There are no clear peaks in the continua, and three of the five participant groups have greater discrimination at the edge rather than middle of the continuum.

Finally, we can return to the finding from the identification results indicating that the /ɨ/ category is at least partially assimilated into the /i/ category. The

prediction from that result is that we should get higher d' scores near /ɨ/ on the /ɪ/ continuum than on the /i/ continuum for English speakers. We can see that this prediction bears out for all levels of English speakers, whether we consider the 1-step or 2-step discrimination scores.

4.4 Simulation 4: One-Dimensional Continua Between English and Russian Phonemes

4.4.1 Identification Fitting

Fitting the 1-dimensional continua around the edge of the stimulus triangle gives us insight into the differences between perceiving a continuum between two native phonemes and a native and foreign phoneme. I begin by fitting the three continua in the experiment to the model for all groups in the experiment. Table 4.5 shows the full set of model fits for all learner levels and continua. We can see that while the categories found for the /i/ and /ɪ/ phonemes do not change as native English speakers learn Russian when we consider perception along this native continuum. However, the model fitted optimal category for the Russian vowel /ɨ/ changes greatly as we look at different proficiency levels. These changes are evident for both the /i/-/ɨ/ and the /ɪ/-/ɨ/ continua, each connecting a native English vowel to the Russian vowel /ɨ/.

Learner Level	c_1 Mean	c_2 Mean	c_1 Variance Sum	c_2 Variance Sum
Continuum between /i/ and /ɨ/				
	$\mu_{/i/}$	$\mu_{/ɨ/}$	$\sigma^2_{/i/} + \sigma^2_S$	$\sigma^2_{/ɨ/} + \sigma^2_S$
Native English	10	12.08	0.593	0.398
Beginner	10	13.59	4.245	2.589
Intermediate	10	13.21	2.666	1.534
Advanced	10	12.37	1.031	0.662
Native Russian	10	11.70	0.396	0.287
Continuum between /i/ and /ɪ/				
	$\mu_{/i/}$	$\mu_{/ɪ/}$	$\sigma^2_{/i/} + \sigma^2_S$	$\sigma^2_{/ɪ/} + \sigma^2_S$
Native English	10	11.87	0.492	0.344
Beginner	10	11.89	0.570	0.396
Intermediate	10	11.88	0.228	0.192
Advanced	10	11.97	0.415	0.284
Native Russian	10	12.72	1.118	1.008
Continuum between /ɪ/ and /ɨ/				
	$\mu_{/I/}$	$\mu_{/ɨ/}$	$\sigma^2_{/ɪ/} + \sigma^2_S$	$\sigma^2_{/ɨ/} + \sigma^2_S$
Native English	10	12.47	0.412	0.407
Beginner	10	12.17	1.221	0.861
Intermediate	10	12.66	0.448	0.685
Advanced	10	12.60	0.308	0.470
Native Russian	10	11.61	0.350	0.258

Table 4.5: Model fits for all levels and continua for identification data. The means for the first category are arbitrarily set to 10 in the Barks scale for fitting purposes.

4.4.1.1 /i/-/ɪ/ Continuum

The found categories for the /i/-/ɪ/ continuum at all levels of proficiency look basically the same, modulo a slight difference for the intermediate speakers (see Figure 4.7).

Particularly, not only do the variances stay largely the same, but the mean of the category does not change substantially between groups (see Table 4.5). This provides several pieces of information. First, it indicates that the model accurately captures underlying dynamics even when the data are different, collected from different populations at different locations. Second, it reflects that the behavior of these different groups is not very different, indicating that these categories are actually not undergoing a significant change in the process of L2 Russian learning. Third, it gives us a baseline for what the variances of these categories are in relation to other native categories. This can be used as a baseline for judging the accuracy and meaningfulness of results when considering continua with the Russian vowel /ɨ/.

4.4.1.2 /i/-/ɨ/ Continuum

When we consider the /i/-/ɨ/ continuum, we see that there is a strong effect of L2 proficiency on the most likely underlying categories that the model finds (Figure 4.8).

Looking at the progression from beginner to advanced learners we see two parameters changing that are meaningful for these simulations. First, the mean for the /ɨ/ category starts out fairly outside the expected area for this phoneme and

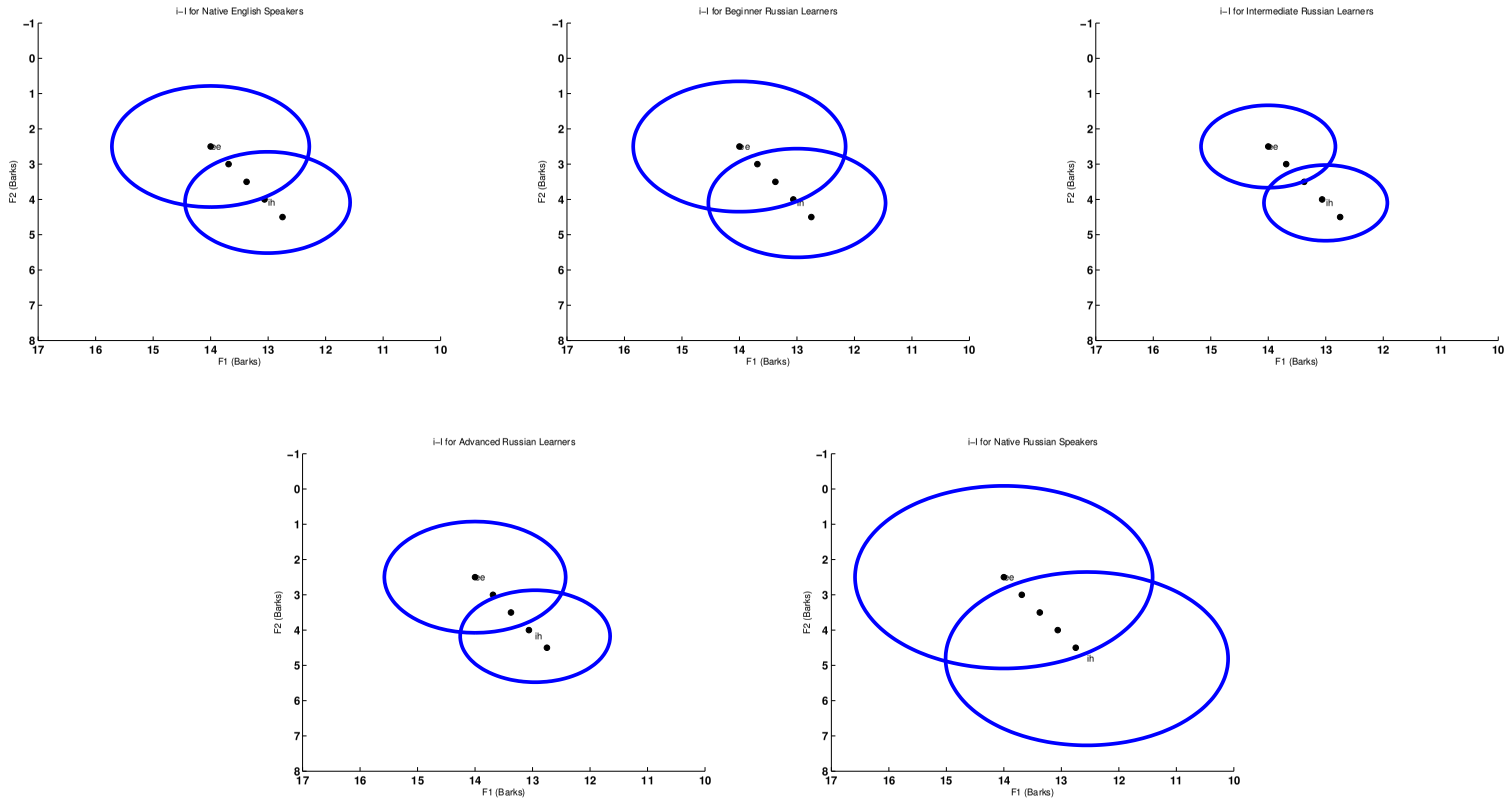


Figure 4.7: Categories found for all groups of participants for the /i/-/ɪ/ continuum. a) Native English speakers, b) Beginner Russian learners, c) Intermediate learners, d) Advanced learners, e) Native Russian speakers

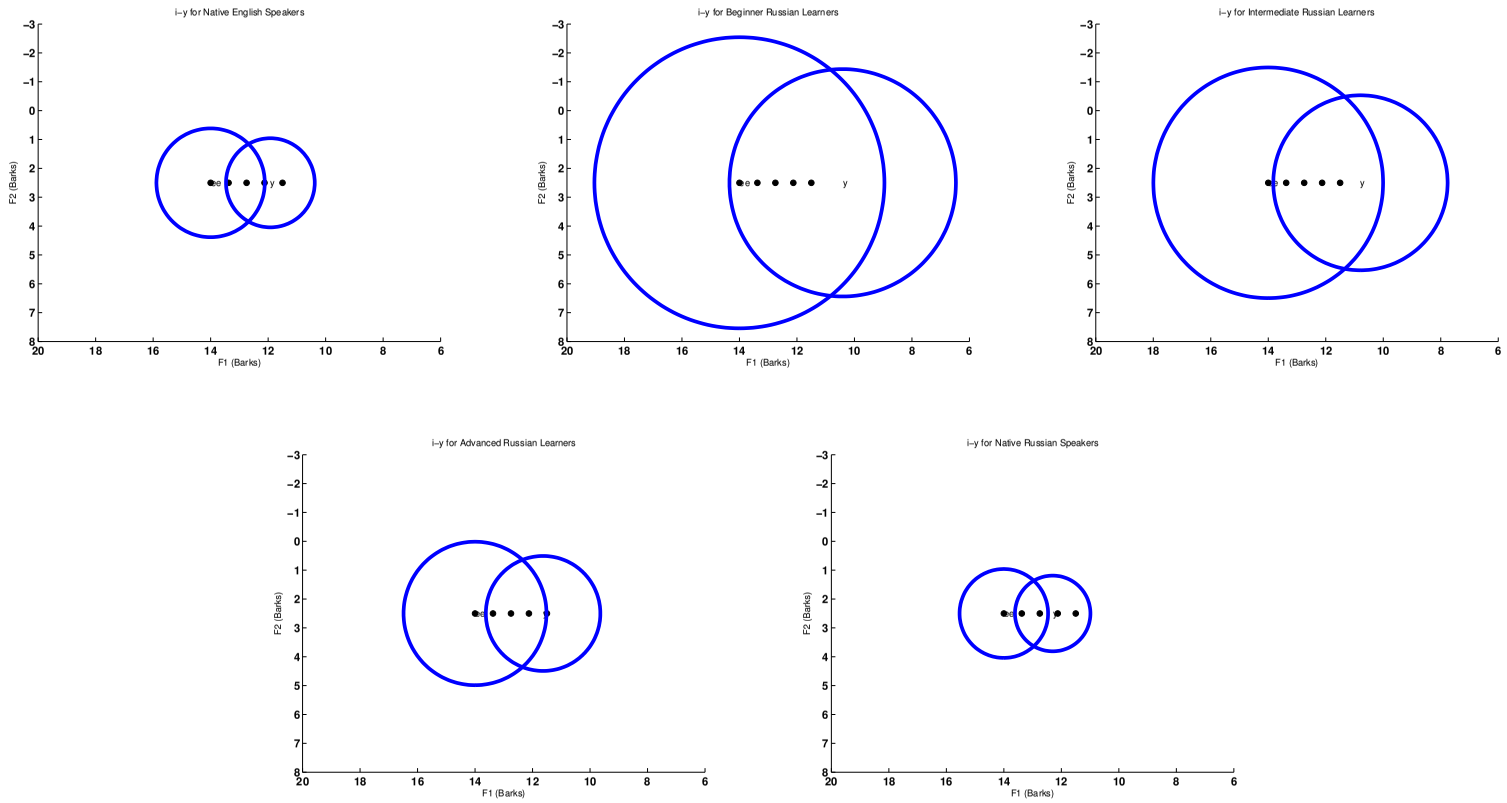


Figure 4.8: Categories found for all groups of participants for the /i/-/ɨ/ continuum. a) Native English speakers, b) Beginner Russian learners, c) Intermediate learners, d) Advanced learners, e) Native Russian speakers

slowly moved up on the F2 measure as people become more and more proficient. Second, we see the variance for this same category becoming smaller and smaller. Together, these suggest that as L2 proficiency increases, people are gaining a more accurate representation of the category structure. If these simulations are taken to correlate with actual mental representation for the categories, then it suggests that it might actually be the ability to accurately represent these two parameters that is responsible for lack of proficiency when starting to learn a foreign language.

Also, it is interesting that the category variances for the /i/ category found in these simulations are very close to those found in the /i/-/ɪ/ simulations. This gives some more credence to these being meaningful measures. There is another interesting finding here. That is that the native English speakers seem to be performing closer to the native Russian speakers than any of the L2 learner groups. They show a tighter variance around the category centers than any of the learner groups, and an extracted category mean for the Russian /ɨ/ vowel that is closest to the native Russian speakers, located slightly toward the /i/ vowel from the continuum endpoint. Of course, these participants do not have any real representation of the /ɨ/ category. It is then surprising that they are behaving as if they do. I hypothesize that the native English speakers are treating this third category as an "Other" category and judging tokens as belonging to it that are bad exemplars of the two native English ones. By this metric, their behavior would reflect proficient learners since the only stimuli that they would end up treating as "Other" would be those that are located where the Russian category /ɨ/ is. However, this ability to accurately do identification of the categories is unlikely to carry over into accurate

discrimination and production of the category since there is no actual representation of the category to rely on. Aside from treating the third category as an "Other" category, the naive learners might also be influenced by the training phase of the experiment. Since they have not been exposed to the category in the past, their representation for it may be formed based on the 5 tokens that they got to hear of the best exemplar. Their category might then be represented as consisting of this one prototype with no variance. If this is the case, we would expect strong warping just as we would with native speakers who have a fully formed category. I don't believe, however, that this training would have an effect on learners with actual experience with an L2. After a year or more of Russian learning, hearing 5 instances should not affect their representation for the Russian category with which they would already be familiar.

Coming back to the location of the category mean for the /ɨ/ category, we can revisit the prediction about the assimilation of /ɨ/ into the /i/ category. Because listeners are likely making this equivalence classification, the prediction would be that they would push the category center for the new category farther away to keep them separate. This is precisely what we see when the beginner learners seem to have a new /ɨ/ category much lower on the F2 scale than expected. As they gain more familiarity with the vowels and become better able to make the necessary distinctions, the category shifts to a more native position.

4.4.1.3 /ɪ/-/ɨ/ Continuum

Finally I consider the third side of the triangle, the /ɪ/-/ɨ/ continuum. Here again we see a strong effect of L2 proficiency on the model fit categories (Figure 4.9). As we consider the categories from beginner to advanced learner, we again see that the found categories reflect a tightening of the variance around the category mean for the Russian category /ɨ/. However, unlike the previous continuum, here the mean of the category does not change much and remains roughly in the actual location, and close to that found in the other simulation. Again we see the native speakers performing closer to the intermediate and advanced learners of Russian than the beginner learners. Again, I believe this has to do with the treatment of the third category in the experiment as a catch-all other category rather than reflecting any attempt to actually classify sounds into a meaningful third category.

It is interesting that on this continuum only the variance seems to need to converge to the final category while the mean is fairly accurate from the beginning. This could reflect the fact that as there is no equivalence classification between /ɪ/ and /ɨ/ for native English speakers, they are able to fairly accurately set the mean as they form the new category.

4.4.1.4 Identification Discussion

Together, these simulations show that the model is capturing meaningful changes in underlying language categories as people progress through stages of L2 learning. For the continuum between two native English vowels, /i/ and /ɪ/, the

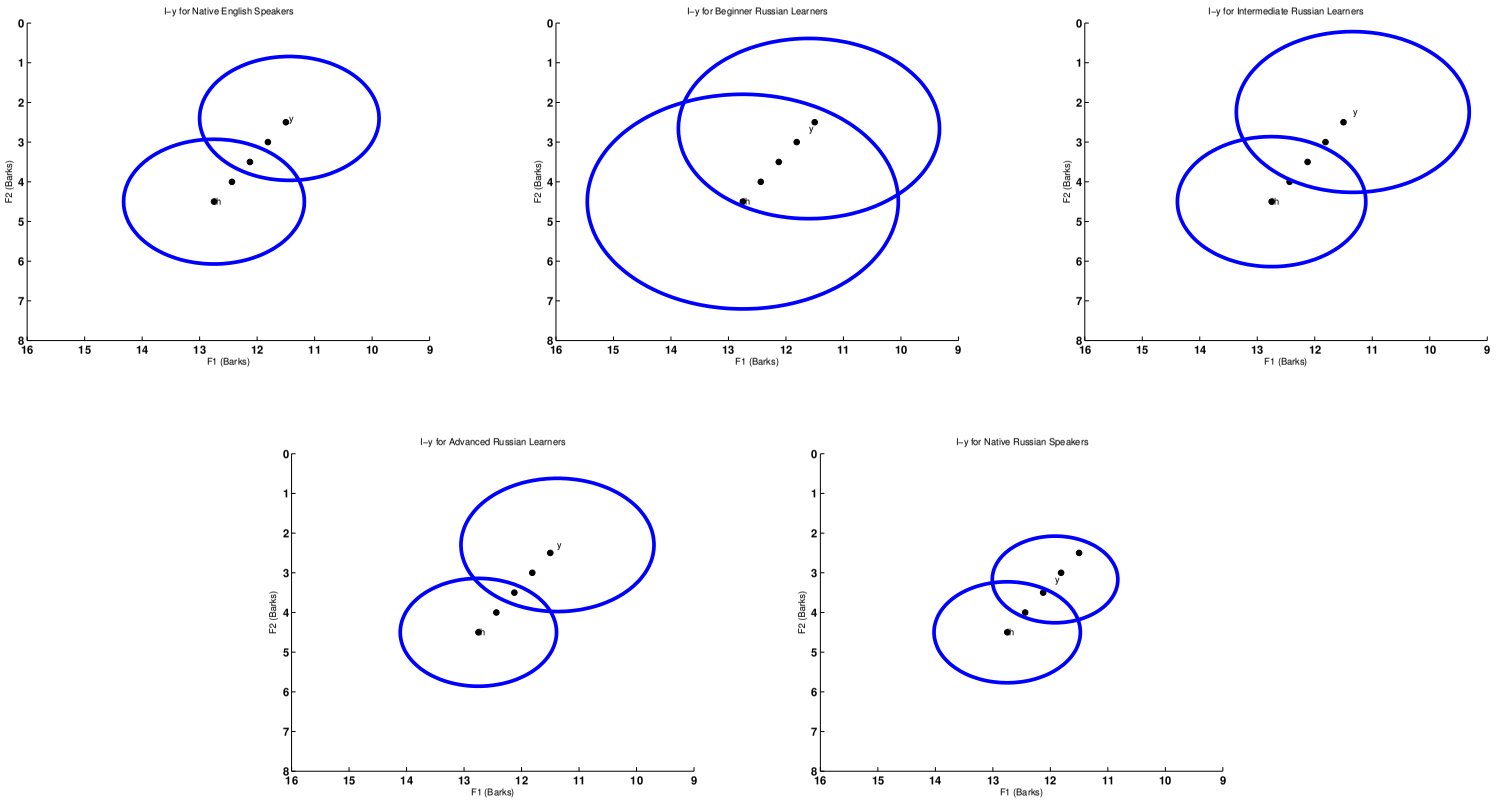


Figure 4.9: Categories found for all groups of participants for the /ɪ/-/i/ continuum. a) Native English speakers, b) Beginner Russian learners, c) Intermediate learners, d) Advanced learners, e) Native Russian speakers

category parameters remain constant for all experimental groups. This is predicted since we do not expect the native English categories to change greatly with Russian learning, at least not for anybody less than a fully proficient L2 speaker. According to the Speech Learning Model (J. E. Flege, 1995), a native category may shift in response to a newly formed L2 category, especially when they are close together, so it would be possible. However, even the most advanced learners in our experiment are only 4th year students, so we would not expect this effect with them. When we consider both the continua between one of the two English vowels and the Russian vowel /ɨ/, we see a very different pattern. Here, we see changes in all category parameters, including both the mean and variance. In general, the categories become more concentrated with a more accurate mean. I interpret this as the L2 learners getting a better representation of the category after being exposed to it in more contexts and many more times as learning progresses. It would make sense that as they increase exposure they would be able to make a much more accurate prediction of what sounds in the category should sound like, hence reflecting these tighter categories.

One possibility for the strange results we are seeing with the beginner learners of Russian is that they do not have the same prior probabilities for the categories as native English speakers or advanced learners of Russian. Normally when running simulations over experimental data I assume an equal prior for all categories, since I do not expect a participant to prefer one category over another or expect it to appear more often in the course of the experiment. However, for a beginner learner of a new language it is not unreasonable for them to have a lower prior probabil-

ity for the new category. In order to investigate this possibility, I reran many of the simulations with different settings for priors to see how the found categories would change as a result. Simulations for beginner learners of Russian for all three 1-D continua for both 50-50 and 75-25 priors for the two categories can be seen in Figure 4.6. What I found was rather interesting. For the /i/-/ɪ/ continuum, changing the prior for the categories had a very minute effect on the categories the model found. This is strange since there is a trade-off relation between parameter values, particularly the prior of the category and the category means and variances. Hence, we should see something similar when we change the prior independent of the continuum. However, this suggests that every parameter has an independent effect beyond the trade-off relationships with other parameters. Hence, while we do see a small shift in the mean and variance optimal fits by changing the prior, it is a much smaller effect since for these categories the other parameters reflect native perception. However, for both the /i/-/ɨ/ and the /ɪ/-/ɨ/ continua the effect was substantial. As we increase the prior on the native English category, the variance of the Russian category decreases and the mean moves closer to the English one. This is most pronounced for the beginner learners, who have a much more reasonable set of categories found with the 75-25 prior as compared to either the earlier 50-50 or the extreme 90-10 (not shown for space considerations). Hence, it is possible that this does in fact relate to how these learners are behaving in the experiment. As I already observed, it is likely that the different parameters work in tandem independently, and not only in a tradeoff relationship. Hence, while this suggests that the category priors should indeed not always be set to 50-50, it does not invalidate

the discussion earlier of the role of category variance and mean as an explanation of what goes on in early to advanced L2 language learners.

Finally I want to revisit the assimilation of the /ɨ/ category into the /i/ category. We saw in the fits for the /i/-/ɨ/ continuum that the category mean for /ɨ/ overshot the expected value by a substantial distance. However, in the simulations for the /ɪ/-/ɨ/ continuum the mean was where we would expect. We need to reconcile these findings since there can only be one actual center for the new category. This is one limitation of modeling the three continua separately rather than doing a full multidimensional fit. While the two continua are not orthogonal to each other, we can still imagine a mean for the category that is slightly lower on the F2 continuum and slightly higher on the F1 continuum than expected, so that when we project it down to the two axis we work with we get precisely the results we see. However, this is something that will have to be left for future work.

4.4.2 Discrimination Fitting

Based on the identification fits along the three vowel continua I can also apply the discrimination fitting procedure from Chapter 3 to extract optimal fits for the underlying category variances and perceptual noise variances. Table 4.7 shows the model output for the noise variance, underlying category variances, and the τ ratios that the model finds to be optimal parameters to account for the behavioral data.

For the native English continuum between /i/ and /ɪ/ we have a reasonable set of findings. All four native English-speaking groups were found to have similar

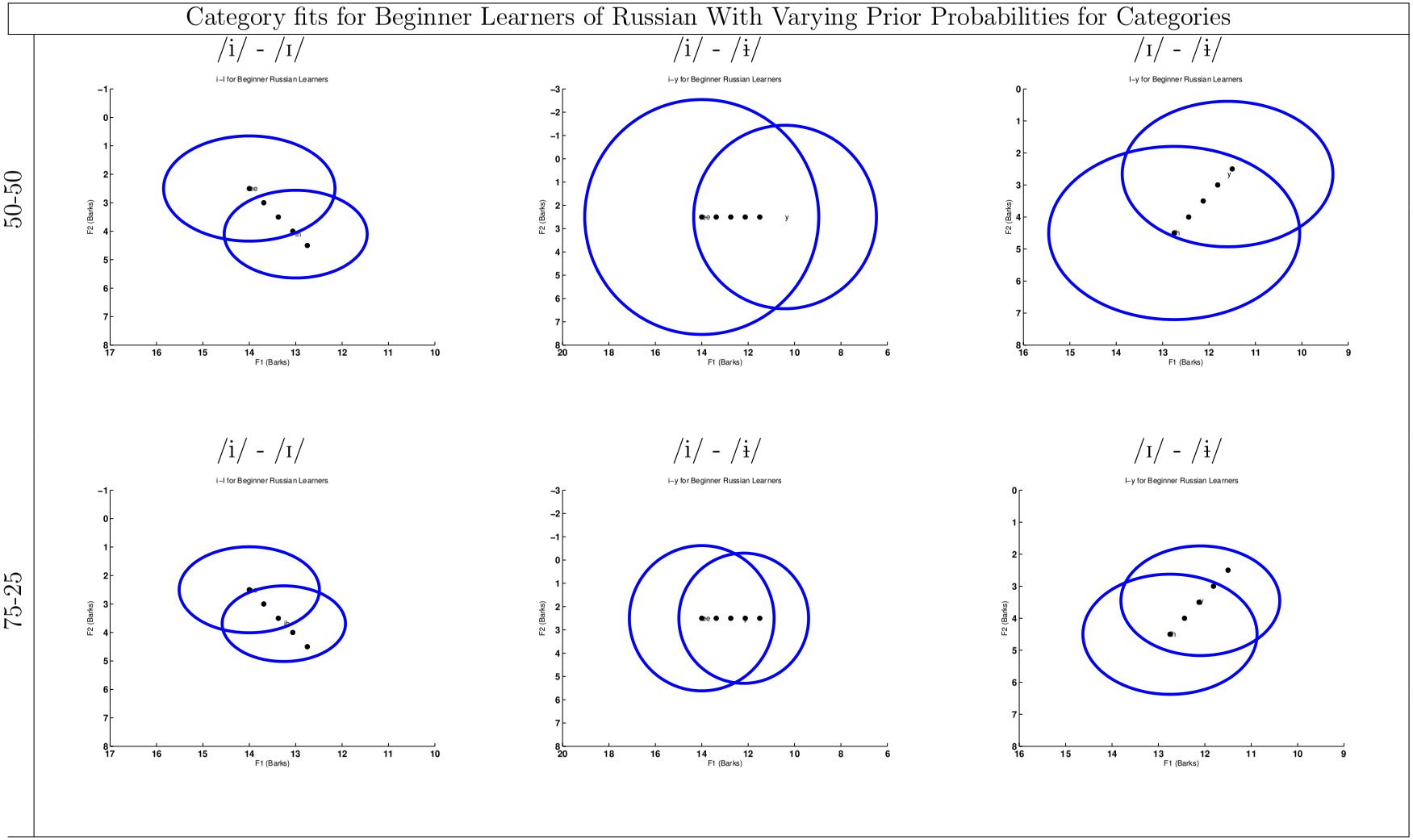


Table 4.6: 50-50 and 75-25 beginner category fits for 1-dimensional simulations.

Learner Level	S Var	c1 Var	c2 Var	c1 Ratio	c2 Ratio
Continuum between /i/ and /ɨ/					
	σ_S^2	$\sigma_{/\text{i}/}^2$	$\sigma_{/\text{ɨ}/}^2$	$\tau_{/\text{i}/}$	$\tau_{/\text{ɨ}/}$
Native English	0.261	0.332	0.137	1.27	0.52
Beginner	4.288	-0.043	-1.698	-0.01	-0.4
Intermediate	2.473	0.193	-0.939	0.08	-0.38
Advanced	0.230	0.801	0.432	3.48	1.87
Native Russian	0.182	0.214	0.105	1.18	0.58
Continuum between /i/ and /ɪ/					
	σ_S^2	$\sigma_{/\text{i}/}^2$	$\sigma_{/\text{ɪ}/}^2$	$\tau_{/\text{i}/}$	$\tau_{/\text{ɪ}/}$
Native English	0.157	0.335	0.187	2.14	1.19
Beginner	0.281	0.288	0.115	1.03	0.41
Intermediate	0.113	0.115	0.079	1.02	0.70
Advanced	0.138	0.277	0.146	2.02	1.06
Native Russian	0.000	1.118	1.008	1.0 E09	9.0E08
Continuum between /ɪ/ and /ɨ/					
	σ_S^2	$\sigma_{/\text{ɪ}/}^2$	$\sigma_{/\text{ɨ}/}^2$	$\tau_{/\text{ɪ}/}$	$\tau_{/\text{ɨ}/}$
Native English	0.089	0.323	0.317	3.63	3.56
Beginner	0.861	0.359	-0.001	0.42	0.00
Intermediate	0.055	0.393	0.630	7.14	11.45
Advanced	0.065	0.243	0.405	3.72	6.21
Native Russian	0.013	0.338	0.246	26.79	19.50

Table 4.7: Model fits for all levels and continua for discrimination data. The noise variance is set by optimally fitting the model to the discrimination data and all other parameters are extracted based on identification findings.

ratios for the two categories and relatively similar values for the perceptual variance parameter. It is worth noting, however, that the ratio values did not correspond with those found for native language vowel perception in Chapter 3. This may be due to a number of factors, including a different set of vowels being tested, a different number of steps on the continuum, a different experimental setup, or a different method of analysis using d' scores rather than fitting an epsilon for direct modeling of same/different responses. But either way, the values were consistent and reasonable. When we look at the native Russian speakers for the same continuum, they perform very differently from the English speakers. For them, the optimal model fit is with no perceptual noise at all and all variance attributed to the category variance. This would correspond to much more continuous perception, which makes sense since the Russian speakers do not have a second category along this continuum to warp their perception there and we saw from the d' scores that they did not have a definitive peak in discrimination between these vowels.

The values found by the model for the other parameters are not reasonable. Particularly, the best fit for Beginner Russian Learners for both the /i/-/ɨ/ and the /ɪ/-/ɨ/ continua find negative values for at least one of the two meaningful category variances. Or, in other words, it finds that the best fit for the discrimination derivation is a noise variance greater than the sum of variances. Of course, this is pathological in terms of realistic human behavior. But, this is perhaps a good result for the model. As mentioned in the previous section, the d' scores for these other two continua were not as expected from perception by listeners with two categories acting to affect the perceived sounds. It is therefore perhaps encouraging

that given pathological inputs the model finds pathological values for the underlying parameters.

However, there were two other findings from fitting the discrimination data that do seem to shed some light on the perception of the participants in the experiment. The variances and τ values found through the fitting procedure for Russian native speakers for the /i/-/ɨ/ continuum are comparable to those found for all the different native English speakers for the /i/-/ɪ/ continuum. This suggests that listeners perceiving stimuli along a fully native continuum behave similarly independently of their native language. The other interesting finding is that the optimal noise variances for native English speakers in both continua between a native English category and /ɨ/ were much higher than the noise variances found for all other proficiency groups for those continua. First, this is relevant because it might help explain the shallow curves of the identification function if the beginner learners are treating this task as a very noise channel, leading to very non-deterministic identification behavior. Second, it helps explain why even the native categories on these continua were found to have higher variances for beginner learners. In the identification fitting, the categories represent the sum of both the meaningful and noise variances. These higher noise variances are driving both categories up. Meanwhile, at least for the /ɪ/-/ɨ/ continuum, we can see that the fitted underlying category variance for the native English /ɪ/ stayed more or less constant between all native English participant groups.

4.5 Simulation 5: Assessing Model Applicability to L2 Two-Dimensional Stimuli

4.5.1 Extending the Model to Multiple Dimensions

Rather than only looking at the 1-dimensional subsets of the 2-d continuum, we can extend the model in order to fit all the categories simultaneously. For identification, we can replace the mean and variance for the categories with a 2-dimensional array of mean values for F1 and F2 and replace the variance with a covariance matrix. We can replace the unidimensional normal distributions with the multivariate Gaussian form:

$N(\mu, \Sigma) = \frac{1}{(2\pi)^{\frac{k}{2}}|\Sigma|^{\frac{1}{2}}} exp\left[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right]$

and then solve to get the identification function:

$$p(c_1|S) = \frac{1}{1 + \sum_{c \neq c1} \sqrt{\frac{|\Sigma_{c1\_S}|}{|\Sigma_{c\_S}|}} exp\left[\frac{(S-\mu_{c1})^T\Sigma_{c1\_S}^{-1}(S-\mu_{c1}) - (S-\mu_c)^T\Sigma_{c\_S}^{-1}(S-\mu_c)}{2}\right]} \tag{4.1}$$

where $\Sigma_{c\_S} = \Sigma_c + \Sigma_S$ and $\Sigma_{c1\_S} = \Sigma_{c1} + \Sigma_S$. (See Appendix D for a full derivation.)

For the discrimination function, we need to calculate a new equation for $p(T|S)$. Using the multidimensional form of the Gaussian we arrive at the representation in Equation 4.2 (See Appendix E for a full derivation).

$$E[T|S] = \sum_c p(c|S)(\Sigma_S^{-1} + \Sigma_c^{-1})^{-1}(\Sigma_S^{-1}S + \Sigma_c^{-1}\mu_c) \tag{4.2}$$

The actual procedure for fitting the identification and discrimination data is overall the same as with the 1-dimensional simulations. We need to fix at least some of the parameters in order to constrain the solution space. Otherwise there would be too many different parameter values that would lead to the same prediction. Once we set these variables, we can then fit the identification data in order to get the means and variances for all the involved categories.

4.5.2 Attempted Model Fitting

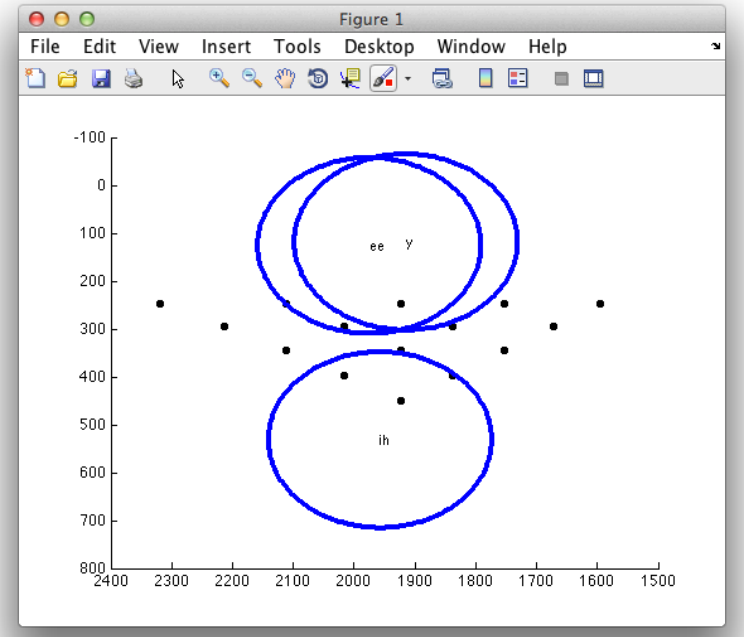


Figure 4.10: Categories found for advanced Russian learners when variances were fixed and model was used to fit the optimal means for the three categories

Following up on the 1-dimensional results, I attempted to use the model to directly fit the multidimensional data. Ideally, we would be able to recover all

three categories directly from the data. However, because there would be too many parameters to fit with the data we have, some had to be set based on production data or other assumptions. With this simplification, I attempted to use the model to capture the full range of human behavior. Unfortunately, the model is not able to capture human behavior when we consider the full 2-dimensional continuum of identification and discrimination scores. I ran four types of simulations: 1) fix the /i/ mean and variance and let the model fit all other parameters, 2) fix the equal variances for all categories and let the model determine the optimal category means, 3) set all category means based on human productions and fit the model to derive category variances, and 4) fix the means and variances for the English /i/ and /ɪ/ categories and fit the model to get the mean and category for the learned Russian /ɨ/ category. However in all these cases the categories based on optimally fitting the behavioral results provided categories that were unreasonable. This indicates that what people are doing is not being fully captured by this model, and hence just a biasing of the acoustic values towards the categories is not the whole story for what people are doing in this experiment.

For illustrative purposes here are some typical results that are obtained when the model is run on the 2-dimensional data. Since these are meant to convey the problem with using the model to fix the categories together, I only include one representative set of results, with all figures for data from the advanced learners of Russian.

In the first simulation I fixed all the variances and ran the model simulation in order to fit the means for the three categories (Figure 4.10). In the second

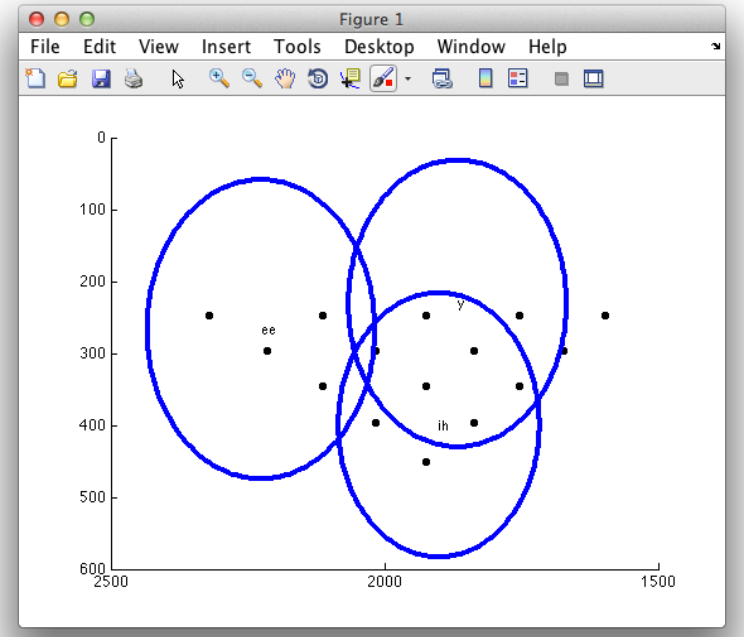


Figure 4.11: Categories found for advanced Russian learners when categories /i/ and /ɪ/ were fixed and model was used to fit the mean and variance for the /ɨ/ category

simulation, I held the means and variances for the English /i/ and /ɪ/ categories constant and fit the model to the data to extract the mean and variance for the Russian /ɨ/ category. In this simulation the outcome was different depending on the initial conditions passed to the minimization function. This indicated that it was prone to getting stuck in local minima. Figure 4.11 shows the found categories given a starting point of $F1 = 200$, $F2 = 1550$, and $\sigma_c^2 = 60^2$. Figure 4.12 shows the found categories given a starting point of $F1 = 200$, $F2 = 1550$, and $\sigma_c^2 = 20^2$. In the final simulation, I fixed the means for the three categories and ran the simulation to extract the optimal variances. This simulation is a bit artificial since there is no reason to expect all learner groups to have an accurate representation of the mean of

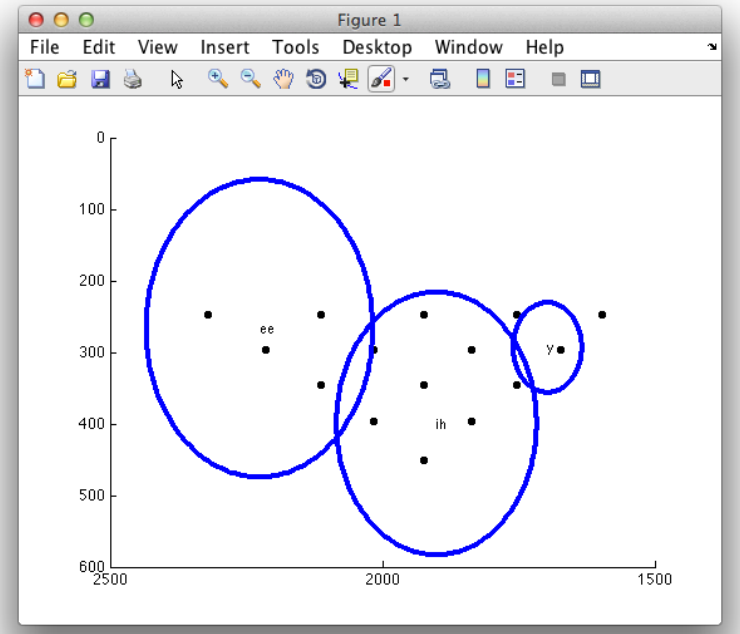


Figure 4.12: Categories found for advanced Russian learners when categories /i/ and /ɪ/ were fixed and model was used to fit the mean and variance for the /ɨ/ category

the newly learned Russian /ɨ/ category. However, if the variances are the dominant category parameter changing over the course of learning, then this simulation would be appropriate. Figure 4.13 shows the result of this simulation.

These strange model fits have two possible causes. First, it might have to do with the fact that I had people work with categories from two different languages. It is possible that representations for a new category in a learned language are not actually represented the same way as native categories. A test of this idea would be to experiment with fully bilingual people, whether through childhood learning or adult acquisition, to see if their results still cannot be captured by this model. However, the problem may have to be more generic, having to do with

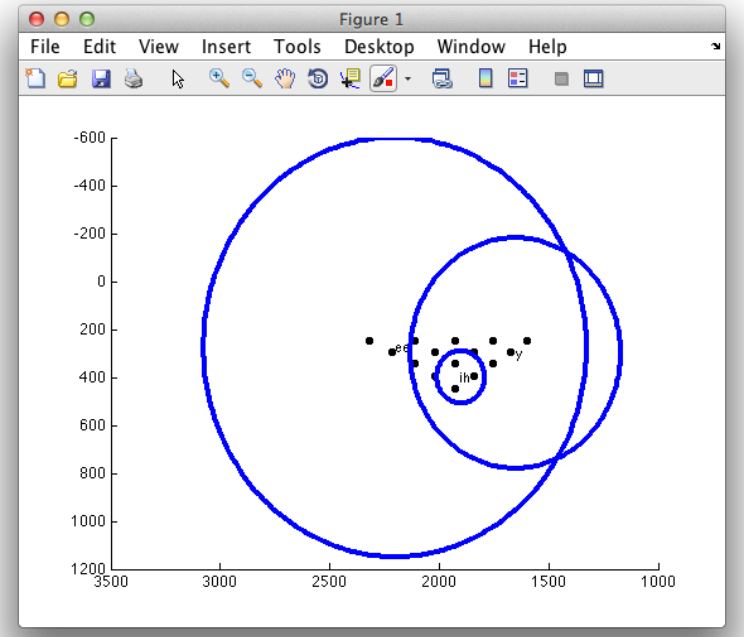


Figure 4.13: Categories found for advanced Russian learners when category means were fixed and model was used to fit the optimal circular symmetrical variances for the three categories

multidimensional perception in general, even if all categories are native ones. In order to examine this possibility, a similar experiment can be run on native English speakers where they are asked to identify vowels in English only with explicit labels. Ideally we would be able to find a set of phonemes that also form a triangular pattern like in this experiment, but this is not a common formation in a single language, especially if we wanted to try to get the vowels to be close together. One possibility would be to do Russian speakers with /i/, /ɨ/, and /e/. If we wanted to do English we could try to have them identify /o/, /ɔ/, and /u/. However, these are not easily modeled as just varying in F1/F2 space and do not form a neat triangle.

4.6 L2 Discussion

In this chapter, I showed how employing the Bayesian model provides insight into the nature of language change when it comes to specific parameters of acquired categories. Also, I uncovered some weaknesses in the model when it comes to modeling complex 2-dimensional data and trying to cover discrimination data for categories that have unequal status for the listener. An important takeaway is that perhaps there is more to examining categorical effects in L2 acquisition than modeling the new categories as being on equal footing to the native categories people bring to the table. However, despite the inadequacy of the model, we are able to learn some important things about category changes over the course of learning.

The identification simulations for individual continua around the stimuli triangle gave us insight into the equivalence classification of the Russian vowel /ɨ/ and also about variance and category mean changes over time. Particularly, English speakers seem to be assimilating tokens of the /ɨ/ category into their native /i/ category. This leads the learners to posit a center for the /ɨ/ category that is farther away from /i/ than it should be for proficient or native Russian speakers. When we examine the /ɪ/-/ɨ/ continuum we do not see this effect since there is no assimilation of /ɨ/ vowels into the /ɪ/ category. Hence, the model optimal mean for this continuum is closer to the native English vowel. In order to examine the true mean of the category we would want to have a full 2-D model fit to the data. Unfortunately, my simulations showed that this is not possible for the current experiment. We were also able to see what happens with variance for the category

over time. Initially, early learners posit a very large category variance (or rather sum of category and noise variance), reflecting their lack of detailed knowledge of how the category is distributed. Over time, the optimal model variance shrinks to reflect that of a native Russian speakers.

Attempts to pull apart the contribution of noise and meaningful variance for the various 1-dimensional continua were not very fruitful. There are two primary reasons for this, one experimental, and one model based. The model treats all categories as equally robust and consisting of the same representation: mean, variance, prior probability. However, it is possible that when we begin to learn a new language we do not immediately represent new sounds the same way we represent native sounds. That is not to say that we just have a different mean and variance, that is given, if we have these parameter values at all. But rather, the possibility that our representation for the new categories consists of some other set of parameters than mean and variance. Given that adults learn more by explicit learning than children, it is possible that the second language category is being represented as a set of exemplars rather than an actual distribution. Also, even if we have these parameters for the new category, there might be additional representational differences such as degree of confidence that in the parameters or even in the existence of the category itself. Further, at some point in the learning phase, the learner has to posit the existence of a new category. Hence, the learning problem is not really just about setting the parameters, but even discovering the need to create the new category to begin with. This I discuss further in Section 5.2.2 when I consider how a learner could compare two possible models with a different number of categories.

Then there is the experimental issue. Whereas simulations in chapter 3 were based on continua of 9, 11, and two sets of 13 tokens, our continua here only contain 5 stimuli. That means we only have 4 one-step and 3 two-step d' scores to base our modeling off. Hence, even for discrimination scores that are not pathological, the fitting procedure is not very reliable. The problem is not just that with 4 points we don't get as good a fit to the curves. It is also that the experimental data is less sensitive and less detailed to begin with. With these continua, we may not be capturing the finer-grained details of participants' actual perception.

There is a possibility that all our Russian learners are at too early a stage in their L2 development to have meaningful discrimination processes. In general, the identification of phonemes can get more native-like before discrimination can. This is because identification can be made based on acoustic measurements and heuristics based on context, while discrimination will only change as warping in the perceptual space changes. This perceptual warping cannot change until there is a new factor at play in doing the warping. In other words, until a category is actually formed, the perceptual space will not change. This may be why we see meaningful results from identification fitting but strange patterns in discrimination. There is other evidence that suggests that identification and discrimination are changing separately during acquisition. In a separate analysis of these findings, Kronrod, Barrios, Winn, Idsardi, and Feldman (2014) showed that the correlation of identification patterns and discrimination performance as they relate to predictions made by the Perceptual Assimilation Model change with Russian acquisition. Particularly, this was true along the native-nonnative continua /i/-/ɨ/ and /ɪ/-/ɪ/ but not for the /i/-/ɪ/ continuum.

In order to asses the validity of this possibility, I would want to conduct a follow-up experiment with highly proficient L2 learners of Russian, likely with participants who have lived in a Russian-language environment for years. Comparing these very early learners to the highly proficient learners as well as native speakers may speak more directly as to the ability of the model to capture L2 effects.

Chapter 5: Discussion and Conclusion

5.1 Summary

The main goal of this dissertation was to establish a computational model that can be used for analyzing categorical effects in both native language and L2 acquisition settings. Specifically, I focused on speech perception at the level of individual phonemes with the aim of showing that these effects are the result of an optimal speech recognition procedure and do not depend on any particular cue representation, processing mode, or language context. I went about this asking two key questions:

- Can we capture categorical effects found via behavioral studies for various phonemes in native language perception with a unified model?
- Can the same model that captures L1 effects be used to examine L2 categorical effects and provide a formal quantitative framework to existing theories of L2 speech perception?

To address these questions, I extended a Bayesian model of speech perception and used both existing and self-collected empirical data in order to find optimal fits for the model to this data. In L1 simulations, the goal was to show that the

model captures static states of perception by looking at native speakers with existing phoneme categories. In L2 simulations the goal was to use the model to examine how particular parameters of the language categories change over time as L2 proficiency increases.

The model that I use is a computational Bayesian model that treats the act of speech perception as an act of optimal inference. At any given point, the listener is trying to infer the speaker's intended target productions. They do this by separating out variability in the speech signal into a meaningful categorical variance and perceptual noise variance. Then, depending on the relative contribution of the two variances, they bias the perceived token from the actual acoustics of the token toward category centers to account for noise in the perception channel. As noise variance increases, listeners are biased toward their category centers. As noise variance decreases and meaningful variance increases, listeners pay more attention to fine acoustic details and perceive stimuli more veridically.

I call this ratio of variances τ. When we fit the model to behavioral identification and discrimination data from various phonemes, the corresponding τ value tells us where on the continuum of degree of categorical effects the particular phoneme falls. In past research, vowels have been found to exhibit very little categorical effects, while stop consonants were shown to exhibit very strong effects. In my model, these varying findings correspond to a different setting of this τ value. In this way, the model accurately captures native language categorical effects in the perception of vowels, stop consonants, fricatives, and nasal consonants within a unified computational model. The found τ values fall on a continuum between 0 and 7, with values

from low to high of 0.02, 0.75, 1.035, 1.86, 1.93, 3.12 and 6.69 for voiced stop /b/, alveolar nasal /n/, bilabial nasal /m/, sibilant fricative /ʃ/, sibilant fricative /s/, voiceless stop consonant /p/, and for vowels on the /i/-/e/ continuum, respectively. Beyond providing a measure of degree of categorical effects with the τ measure, we can automatically extract the means and variances of the categories, which in all cases correspond closely with those we have from production studies.

An important aspect of this unified model is that it is independent of the cues that are used to perceive a particular phoneme and the neural implementation of how these phonemes are perceived. The phonemes I work with vary by cues including F1 and F2 static values for vowels, VOTs measured in ms between closure release and vocal fold vibration for oral stop consonants, the location of two central frication frequencies for sibilant fricatives, and F2 onset transitions for nasal stop consonants. This suggests that these categorical effects come out from some higher and more domain-general categorical effects engine. Perhaps one that is not even speech specific, but rather used for all manner of classification and differentiation of stimuli in situations where there is a natural grouping that underlies the stimuli presented.

Once I established that the model is adequate for capturing a variety of categorical effects in L1 perception, I addressed the question of how this model may be used in order to investigate changes in phoneme perception as L2 learners progress through various stages of fluency in a new language. I conduct this investigation with the assumption that we can model speech perception behavior as originating from the same process of optimal inference independent of how well developed any

given phonetic category is. This assumption may not hold up, as I will discuss in the model limitation section below. But, given this assumption, I collected behavioral data from English-speaking learners of Russian and used the model to find optimal parameters for means and variances of both native English categories and a newly learned Russian vowel category.

I chose vowels to conduct this initial experiment in L2 acquisition because of both the ease of rendering them for the experiment, and the straightforward nature of discussing multi-category perceptual warping, since they are all located in a single F1-F2 space. Of course these vowels do differ on other dimensions such as duration and they are they have diphthong structures, which complicated the stimuli creation phase, but once that was complete the primary analysis could be conducted along the two formant dimensions. In my experiment I considered the high F2, low F1, quadrant of standard vowel space, covering the native English vowels /i/ and /ɪ/ as well as the Russian high central unrounded vowel /ɨ/. I then collected goodness, identification, and discrimination data from three levels of Russian learners and two control groups: monolingual English and native Russian speakers. English monolingual speakers provide a snapshot of behavior before any learning of an L2 occurs. The native Russian speakers gave me a target for ideal perception of the Russian vowel category. The beginner, intermediate, and advanced learner groups were the three critical conditions to see what changes in terms of category parameters as learning progresses.

Contrary to my original plans, I found it impossible to fit the model to either the identification or the discrimination data by considering the multidimensional

data all at once. However, I was able to use the model to assess changes in perception along three 1-dimensional continua that constituted a subset of the vowel space in my experiments. These continua were /i/-/ɪ/, /i/-/ɨ/, and /ɪ/-/ɨ/. The results of the goodness experiment confirmed that participants identified all endpoints appropriately and showed variable within-category goodness ratings which quickly lowered as we moved away from the corner stimuli. From the raw identification results I was able to ascertain that English speakers were assimilating tokens of /ɨ/ into the /i/ category to a high degree, with no assimilation to the /ɪ/ category. Fitting the model to the behavioral identification data further corroborated this fact and provided some interesting insights into parametric change over the course of learning. On the /i/-/ɨ/ continuum, the mean for the Russian category /ɨ/ was set past where it should be by beginner learners of Russian, effectively overshooting the target. This conforms with predictions from the Speech Learning Model (J. E. Flege, 1995) that states that for contrasts in an L2 that are too close to a category in an L1, subjects will often dissimilate the categories by pushing them apart to make perception and production easier. As we go from beginner to advanced learners, the optimal center for the category slowly moves toward the target value at the endpoint of the stimuli triangle. When we look instead at the model fit for the /ɪ/-/ɨ/ category, we find the category center fairly accurately set from the beginning. For both of the continua, beginner learners set a very high variance for the new category, and slowly reduce it toward the target value as learning progresses. The advanced learners seem to be identifying the vowels similarly to the native Russian speakers. Based on this I propose that while the category is set based on a combination of

raw input and the equivalence classification with the L1, the variance is initially set high in order to ensure capturing all the necessary tokens, and the category is subsequently reduced to proper size once the learner is confident in the distribution.

While the advanced learners seem to behave similarly to native Russian speakers on the identification task, when we examine discrimination along these continua, my advanced learners do not look so advanced anymore. At least in the sense that their discrimination performance does not approach that which we would expect of a person with all three categories fully in their category inventory. As expected, all English speakers had nice discrimination patterns with peaks at the category boundary for the native English /i/-/ɪ/ continuum. On the /ɪ/-/ɨ/ continuum, English speakers showed a peak and markedly higher discrimination a the /ɨ/ end than the /ɪ/ end. This suggests that perception around the non-native vowel is higher because it is not robust enough to warp the perceptual space around it in the same way. When we consider the performance of the English speakers learning Russian on the /i/-/ɨ/ continuum, the discrimination scores are somewhat pathological. There are no clear peaks, except for the intermediate group, and worse discrimination at the /ɨ/ end. This is the opposite of what I would expect given that this category should have less of a perceptual warping effect on the surrounding area, akin to the /ɪ/ category for Russians. This may again be due to the assimilation of tokens from the /ɨ/ category into /i/. If these tokens are all being assimilated as "bad versions of /i/", then this could explain the very low discrimination scores, as some of the trials would just be viewed as equally bad versions of the same category.

5.2 Unified Approach to L1 and L2 Results

In chapter 3, my approach to examining categorical effects followed a classic way of framing the question. I examined identification and discrimination curves and discussed what underlying computational goal would lead to such behavior. The main empirical contribution was providing a means of applying the same model to the different phonemes in order to explain this data. In chapter 4, my approach to considering categorical effects in L2 perception was different from classic SLA literature. I did not frame issues in L2 perception as a consequence of building blocks such as cues and features that might or might not exist in the L1. Also, I did not examine directly the relationship of equivalence classification of pairs of L2 sounds with categories in the L1 and the consequence for discrimination ability, though I did borrow some thoughts from these predictions in my analysis of some contrasts for which I got unexpected results. Instead, I considered categorical effects in L2 perception as derived from the same process of optimal inference that I used to consider effects in L1. To the extent that this approach works, it is an exciting way of merging the study of categories and perception across two disciplines. This approach to L2 perception immediately gives us a way of making formal quantitative predictions for behavior, gives us insight to more fine tuned behavior along a continuum of speech tokens, and formalizes some predictions made by both the PAM and SLM. And, of course, this approach does take care of some of the shortcomings of SLA models that I mentioned before. I examine perception along a continuum rather than just a 2-way contrast of speech tokens. I formalize a way of computing a similarity met-

ric by calculating the distance between intended target productions. I am able to track changes in actual category parameters over time. The part that is not quite complete is the ability to make a-priori predictions about discrimination behavior based on category structures without relying on known identification data. At the beginner stage, when we assume that people have not learned any new categories, we can make predictions based on native categories, for both identification and discrimination. However, this is not something we can do for intermediate learners. If we stipulated some facts, such as the variance starting large and reducing over the course of learning and the mean being set based on initial equivalence classification (as my findings suggest), it would be possible to predict behavior for intermediate stages of learning too. This, however, is beyond the scope of this dissertation.

5.2.1 L1 vs L2 learning

In this dissertation I discussed L2 learning, as well as initial perception, as being colored by category distributions from the L1. Particularly, the L2 learner starts with certain categories that warp the perceptual space, and then over the course of exposure to data from an L2 eventually forms new categories and adjusts their parameters so that they in turn become robust and start to also warp the perceptual space. This may be very similar to what a child who first comes to the task of phonetic category learning has to do in the L1. We can consider infants coming to the task of phonetic category learning as having a starting point based on certain restrictions determined by the auditory system. This may be innate

discontinuities in perception for stop consonants (Pisoni, 1977; Eimas et al., 1971), or innately specified natural auditory boundaries that partition the auditory space for vowels (Kuhl, 1993b). In addition, at least for vowels, infants show an underlying perceptual bias that leads them to favor vowels at the edges of the F1/F2 vowel space (Polka & Bohn, 2011), leading to additional language-independent biases as they embark on the language learning process. In the first year of life, these infants start to respond differentially to categories that are linguistically relevant in their native language and start to collapse category differences that exists in foreign languages (Werker & Tees, 1984). This process of learning may be similar to the process of L2 learning, especially if we consider the claim that adults may use the same learning procedures as to infants (J. E. Flege, 1995). One process that both a child and adult can use to learn new categories is by positing the existence of a new category that covers the entire phonetic space and then using their experience to set the parameters for the category. This is something akin to setting an "Other" category and using it in the perception task, as I discussed in regards to L2 learners. Of course the learner would then have to decide whether they need this new category or not, which is something I discuss in the following section.

5.2.2 Model Comparison

In the L2 section of this dissertation, I conducted my simulations under the premise that even early learners have a category representation for the new L2 categories, but that the category parameters, particularly the mean and variance,

may not be very accurate initially. Then, over time, the learner would adjust these parameters to approach a native-like state. However, it is likely that at the earliest point in L2 learning, the learner would not have any representation for this category, at least not one similar in structure to the ones in the L1. It is an interesting question to ask at what point we have evidence of a new category coming online for the learner. As I discussed in the previous section, we can also consider the learning problem in the L1. How would a child know how many categories they have in their language. Or, to put it another way, independently of whether we are dealing with children or adults, and whether it is in regards to L1 or L2 learning, how can we determine whether a model with $K + 1$ categories explains perceptual behavior better than a model with K categories (in my L2 work, this would be answering the question of when does the Russian category come online for native English speaking L2 learners). In order to do this, we could fit the behavioral data to both models and extract the most likely parameters for the categories that cause this perception. In the process of finding the optimal parameters, we search through the space of parameters and look for the smallest error term, ϵ, between the model predictions and the observed behavioral data. We can in turn use this error term to compare multiple models that differ by number of categories. If the model with $K + 1$ categories fits the data with an error term smaller than that of the K categories, then we can posit that the listener likely has the extra category and can proceed to examine the fitted parameters. If, on the other hand, the K category model has a better fit, it is possible that the listener has not yet formed the new category, or that it is not robust enough to warp the perceptual space in the

same way as the native categories. However, we can't just use the goodness of fit or the error term to decide directly between the models. All other things being equal, the model with more parameters will allow a closer fit to the data, though it might actually be overfitting the data. To account for this possibility with models that differ in complexity, some sort of penalty for extra parameters is typically added. A common penalty metric in such cases is the Bayes Information Criterion (BIC) (Schwarz, 1978). The BIC penalizes the model for any extra parameters, in effect allowing us to measure the efficiency of the parameterized model in terms of its ability to predict the data. Using this approach, we as the modelers can figure out at what point in development or L2 learning the learner is likely to be performing discrimination in a way that indicates the presence of a new vowel. It is further possible that the learners themselves can use something like a model comparison metric to decide at what point to formally add a category to their phonetic space.

5.3 Theoretical Implications

The model developed in this dissertation can unify and replace previous accounts of categorical perception and the perceptual magnet effect (PME). But how does the current model relate to these two classic accounts? Categorical perception is in effect an efficiency construct such that we can avoid burdensome acoustic processes once we have identified the phoneme. Under the CP hypothesis, people automatically assign category labels to stimuli so that in a discrimination task they are really comparing the labels and not the acoustic measurements. This goes

nicely hand-in-hand with current hierarchal models of phonology processing (Hickok & Poeppel, 2000; McClelland & Elman, 1986; Poeppel, Idsardi, & van Wassenhove, 2008). The PME states that prototypical tokens of vowels exert a pull on surrounding stimuli. Under this hypothesis there is no assignment of labels, but rather a direct influence of category prototypes that cause warping of the phonetic space. My model borrows from both of these classic ideas in the following sense. In my model, the computation of the intended target production involves a weighted average that is weighted by the probability of belonging to either category. However, unlike the CP hypothesis, where discrimination depends on explicit assigning of category labels, there is no need for actual classification to occur for this effect to take place. Rather, the contribution from each category may be captured in the representation of the posterior probability of the target productions implicitly. However, even if the use of the $p(c|S)$ term is a mere convenience at the computational level and does not correspond with a calculation of this value explicitly, it is notable that there is a direct effect traceable to the probability of each underlying category as there is under the CP hypothesis. Considering similarities with the PME, my model directly deals with warping of the perceptual space. However, unlike the PME, where warping is based on proximity to the prototypical phoneme, this model warps space based on classification likelihoods. Also, this model warps toward a mean of a category distribution rather than an exemplar with a high goodness rating.

Moving away from CP and the PME to focus on the present model, we can consider in detail at what level this model helps explain perception. The model is meant to capture speech perception at the computational level (Marr, 1982). It

makes no direct algorithmic or implementation claims, even if the computations involved may suggest certain process-level accounts. This is an important distinction, because stating that a computation involving certain parameters is being solved is not the same as specifying that these parameters are either explicitly represented in memory or derived somehow on the fly from acoustic signals. However, for such a computational approach to guide perception, then these parameters do need to be compatible with the underlying algorithm and implementation. As such, our findings do have implications for models at other levels of analysis. In our model, probabilities of sounds in a listener's underlying categories are employed even when there is no explicit identification task being performed.

There are models that have been proposed utilizing exemplars, prototypes, and neural representations that are all compatible with this computational-level account. One such model is Lacerda's (1995) model of categorical effects as emergent from exemplar-based categorization, which has a direct mathematical link to this Bayesian model, as described by N. H. Feldman et al. (2009). Shi, Griffiths, Feldman, and Sanborn (2010) also provide an implementation of this model in an exemplar-based framework, which is closely related to the neural model proposed by Guenther and Gjaja (1996). In that model, there are more exemplars with a higher level of neural firing at category centers than near category boundaries. This occurs even when a percept is closer to the boundary, simulating a partial weighing of the category center on top of the acoustic signal. There is also a second relevant neural network model that relates Gaussian distributions of speech sounds and neural firing preferences (Vallabha & McClelland, 2007). The perceptual and categorical

level had bidirectional links, allowing the category to not only be influenced by the perception, but to in turn influence the perceived sound. This also leads to a bias toward category centers. While I am not proposing that one or the other model is the accurate representation of how the present Bayesian model is implemented, the links between them are worth further investigating to see if together, they can combine to explain the range of categorical effects on perception at all levels of analysis.

At a neural level, it is possible that different implementations are indeed playing a role in perception of different phonemes. The nature of the cues being perceived makes this likely, as different classes of phonemes make use of different cues for encoding phonetic distinctions. Particularly, vowels primarily employ static spectral cues while consonants primarily employ dynamic spectral and static temporal cues. Correspondingly, studies have found that consonants and vowels are represented differently in the brain both in terms of physical location as well as the associated patterns. Obleser, Leaver, VanMeter, and Rauschecker (2010) found that vowel and consonant categories are discriminated in different regions in the entire superior temporal cortex. While the activation patterns were complex for both stimuli, they were sparsely overlapping and the local patterns were distributed across many regions of the auditory cortex. Perez et al. (2013) found that different coding strategies were used for vowels and consonants. They found vowel discrimination corresponded largely to neural spike counts in certain brain regions, while consonant discrimination correlated to neural spike timing and not to counts if timing information was eliminated. This suggests a difference between temporal or dynamic encoding and

static spectral encoding. Therefore, while by a type of Ockham's razor account we would prefer a more parsimonious solution, I leave the question of whether categorical effects in different types of sounds necessarily require different implementations as an open empirical question. On the basis of existing evidence, however, this question seems likely to be answered by an appeal to different implementations.

Focusing on the theoretical claims related to the L1 work in this dissertation, we can ask what the τ continuum really represents. The τ continuum corresponds to the degree of categorical effects on the perception of various phonemes. At the computational level, this value reflects the ratio of meaningful to noise variance for a given category. Vowels were found to have high meaningful variance and low noise variance, whereas consonant were found to have high noise variance and low meaningful variance. Here I consider the question of what exactly these mean, why one type of variance might be meaningful while another is not, and what epistemological status they have.

In plain terms, meaningful variance should be predictable and may be directly related to the mapping of phonetics to phonology, thereby facilitating the use of fine acoustic detail in the retrieval of phonological representations. Meaningful variability may also serve for extracting indexical information. For example, listeners use vowel variation for identifying speakers, dialects, and accents. Vowel inherent spectral change has been found to be central for differentiating regional variations of American English (Jacewicz & Fox, 2013). The presence of meaningful indexical variability in vowels across multiple languages suggests that vowels may be naturally more prone to develop these slight differences than consonants are, and this is why

we learn to treat the differences as meaningful.

Noise variance, on the other hand, should represent variability in the signal that is a side-effect of the way a sound is produced and properties of the sound that make it prone to misinterpretation. For example, if consonants were inherently more prone to interference because of their short duration or because of the nature of their acoustic characteristics, then their lower degree of meaningful variability would make sense. By keeping the target productions narrow, we prevent perception from being too difficult once noise is added to the signal. While noise variability could be as simple as a sound to be pronounced different ways independent of context by a variety of speakers (i.e., the proneness of a particular phoneme to undergo phonetic shift), it could also be a representation that listeners have of the difficulty or likelihood of wrongful execution on behalf of a particular articulator or the human ability to perceive certain specific cues accurately in an average environment.

Moving on to the L2 work, there is also the theoretical question of shared phonological space between the L1 and L2. The Bayesian model in my dissertation relies on a shared phonological space in order to make predictions about L2 processing. Critically, the newly formed L2 categories must be able to warp the same perceptual space that is being warped by the L1 categories. The only other alternative is that people starting to learn an L2 make a copy of their L1 space as a starting point and then modify it with L2 learning, eventually creating two separate spaces. However, much research on early L2 learning shows that this cannot be the case and we do seem to be working in a shared phonological space (e.g., Chang, 2011).

If we establish that we are working in a shared phonological space, there is then the question of whether people represent their L1 categories in the same way that they do their L2 categories. The fact that we were unable to capture some of the behavior in L2 in a clear manner suggests that maybe the representations are not the same. A simpler question is whether they end up the same once the L2 learner nears full proficiency. This would be a simpler question to answer, but would require follow-up experiments on a new population. A related question to this question of same or different representation is whether adults and children go about figuring out the categories in the language the same way. In this work, I implicitly take up Flege's viewpoint and consider L2 learners to have access to the same learning mechanisms as a child. This does not mean that they are quick learners nor that their representations are accurate or resemble native speakers, but merely indicates that the nature of the representation is similar. Of course a child does not start with a fully specified phonetic space, while the L2 learner starts where their L1 left off, so naturally there are major differences. But if they employ the same mechanisms, then the same types of categorical effects should apply to these new categories as it does to the native ones. Or, more precisely, at the computational level, the categories influence the optimal inference mechanism for intended speaker productions in a similar way, independently of exactly how the category representations are neurally implemented.

5.4 Limitations of the Model

One limitation of the model is that it cannot handle different types of representations for different categories, nor does it contain something like a confidence metric for perception when new categories are not fully formed. Particularly, these limitations relate to L2 perception. While I believe there is no issue with treating L2 categories as sharing the phonological space with L1 categories, and my results showed that we can infer meaningful information from observing early L2 learners, there are parts of behavior that the model cannot capture. One instance is that of identification between /i/ and /ɨ/ for English-speaking beginner learners of Russian. They seem to behave in such a way that there is a random element to their decision making or just some higher level lack of belief in their perception that leads them to select the native category a minimum percent of the time (in this experiment, around 30%). While the model does allow for different priors, and I did show that adjusting the prior leads to a more consistent mean and variance setting, it still cannot fully explain the results of the identification function. The question of representation is trickier. Just because there is a different explicit representation, or a different way of processing something at the neural level, does not mean that at the computational level we can't treat the effects jointly. In fact, in the L1 simulations, we saw that while the cues necessary for identifying different phonemes differ greatly as to how they are perceived, the categorical effects are adequately explained by the unified model. Likewise, in L2 perception, we have a good idea that categorical perception in L1 and L2 is implemented differently at the neural

level (Minagawa-Kawai, Mori, & Sato, 2005). While this does not preclude unifying these effects at the computational level, it is currently a limitation of the model that it cannot fully handle this data.

Another limitation of the model is that it cannot handle a situation where certain phonemes are uncategorizable and exist in an area of phonetic space where no category has a warping effect. In the model, the target production is biased toward every category mean by the size of the noise variance and in proportion to the probability of belonging to that category. Hence, even if a sound is far from any category, it will still be biased toward the nearest category, and equally strongly as other sounds near it. This makes it difficult to capture naive perception of an L2 or even beginner learners of an L2 behavior when the new phoneme is one that would not be readily assimilated into a native category. A possible addition to the model could be a fading factor where the bias toward the category center would slowly fade out proportional to the distance or the distance squared from the center of the category. This would make the model better resemble predictions made by the Perceptual Assimilation Theory (Best, 1994) for phonemes that are marked as unassimilated or uncategorizable.

5.5 Limitations of the Experiments and Simulations

Individual Differences

In this dissertation, all the modeling is based on data aggregated across participants in experiments. However, it is known that speakers of a language exhibit

individual differences for a variety of phonetic categories including vowels (J. Hillenbrand et al., 1995; Peterson & Barney, 1952), fricatives (Newman, Clouse, & Burnham, 2001), stop consonants (Allen, Miller, & DeSteno, 2003; Byrd, 1992; Zue & Laferriere, 1979), and liquids (Espy-Wilson, Boyce, Jackson, Narayanan, & Alwan, 2000; Hashi, Honda, & Westbury, 2003). Looking particularly at stop consonants, Allen et al. (2003) showed that even if they accounted for speaking rate via syllable and vowel duration, there was still a substantial amount of variability in VOT values that was accounted for with individual speaker differences. In further work, Theodore, Miller, and DeSteno (2009) showed that beyond individual differences in VOT production values, the effect of speaking rate on those VOT values was also in turn speaker-dependent, though the same study found that when considering differences in VOT productions for different places of articulation, the individual differences were stable at all locations.

The aforementioned findings tell us that individuals vary in terms of production. But this leaves the question of how they vary in terms of perception. We know that as listeners we adjust to individuals and their speaking rates. With vowels for instance, not only do absolute values of the formants vary between speakers, but also their context-dependent changes and relative position to each other, requiring significant normalization for different speakers (Miller, 1981; Joos, 1948). This adjustment does not only affect what we consider to be the category boundary and variance, but also how well individual members of a category represent that category. Theodore, Myers, and Lomibao (2013) showed that when listeners were trained on two voices with one producing low VOTs and one producing high VOTs, then at

test they substantially shifted their goodness ratings depending on the identify of the speaker they were rating. The effects were strong, shifting goodness ratings at what is typically the center of a voiceless stop consonant category from 6 out of 7 down to 3.5 out of 7. However, in that particular study they did not find a category boundary shift. Even simple changes in rate of speech produce goodness ratings changes that don't only affect perception at the boundary, but extend throughout the entire category (Miller & Volaitis, 1989).

These findings raise some issues that need to be addressed related to the claims and findings of this dissertation. First, if people shift their category representations to account for individual speakers, then how strongly does their own individual variation that comes out via production affect their default processing in terms of assumed category means, variances, and goodness ratings distribution? Second, in terms of the modeling in the dissertation, can the model be easily adapted to capture this individual variability or does this pose a terminal challenge to the findings presented here? While I do not suggest a specific change to the model to account for individual variation, I consider some implications below.

As I made clear above, in this dissertation, all of the modeling that I completed is based on data that was aggregated across participants. This includes both identification and discrimination data. Hence, if the aggregated data misrepresents the degree of categoricity, then the model will fail to capture the true underlying human behavior. Particularly related to the L1 modeling, this could lead to erroneous findings in terms of where individual phonemes are located on the τ continuum. The ratio of variances is derived from the model fitting, which in turn is contingent on

the steepness of the ID curve as well as the peakedness of the discrimination curve. Both of these curves will appear shallower than they should be in the aggregate form if individuals have shifted boundaries and peaks. As a result, the model will find much smaller biases toward category centers than individual listeners actually have, and lead to the phoneme falling higher on the continuum than it otherwise should. This may at least partially explain the finding for the voiceless stop consonant category /p/, which was found to be much higher on the τ continuum than its voiced counterpart in my simulations, as past research showed that speakers vary a lot in terms of their VOT productions and effects of rate of speech for this category.

One way around this problem is to only apply the model to individual data. This makes sense especially if we want to use such a model to allow us to observe where individuals have trouble with a new language to target phonetic exercises as a pedagogical tool. However, if the point is to accurately predict L1 perception, then the logic can get a bit circular. In order to predict individual's perception we need to know their categories, and in order to extract their categories we need to see their perception data. One way out of this is by observing that in the individual differences literature described above, speakers were shown to shift their category parameters consistently across contexts and places of articulation, such that the model may be able to maintain a single representation of a category and just fit an additional individual shift parameter based on a few selected examples of the person's speech. Taking a step back and considering what this model is intended to accomplish, its possible that this is not a problem at all. The model is not really all about prediction, but more about capturing the behavioral data and explaining

the source of categorical effects on perception. The point is not necessarily to predict where a specific person and phoneme fall on the τ continuum, but rather to examine how these differences in categorical effects come to be and how they can all be captured by a unified approach.

When we consider the application to second language learning, we don't have to worry about circular logic. Here, in fact, its perfectly reasonable to collect a full spectrum of L1 data from the listener in order to fit all their L1 categories. With the L1 categories in hand, we can make a-priori predictions for how this particular speaker will behave in an L2 perception task or during the course of L2 learning in comparison with other speakers who may have slightly shifter representations in their L1. In fact, this would be a fruitful area of future research as it can start to bridge the gap between theory and SLA applications.

Also, we can consider the relationship of goodness data to the modeling process. As I outlined above in terms of past research, goodness ratings were shown to be strongly affected based on speaker variation and rate of speech. It is then possible that based on one's own category representation, the default category structures vary in terms of what is considered a good member of the category. While the goodness data is not sued to inform the modeling process, it can serve as a good measure to confirm the reliability of the categories over the course of development. In L1 research, goodness has been correlated in past work with discrimination in work looking at the perceptual magnet effect on vowel perception. While it was made clear that goodness ratings don't necessarily correlate with discrimination for all consonants, they may correlate with the expected variance of a category, giving

us some corroborating evidence for category fits. In L2 work, goodness ratings are used to make predictions for discrimination based on being able to tell apart single category assimilation patterns from category goodness assimilation patterns. Over the course of learning, as category structures change, these patterns also change. We can use the differences in individual's goodness ratings to examine the relative changes in identification and discrimination patterns over time and use them to confirm how well defined categories have become for learners. Some initial work examining goodness, identification, and discrimination data collected in these experiment is reported in Kronrod et al. (2014).

We can also consider how the individual differences in production of categories informs the listener's setting of category parameters. As I mentioned, we can account for shifts in means by modeling a shifting parameter to account for individual differences. While this helps in the eventual perception task and accounting for speaker differences, it doesn't tell us how the categories were learned from the multiple speakers. This individual variability is part of the input for all listeners that they use to form their own category representations. It is then worth considering how we can use this production variability to inform the modeling process. In this dissertation, I use values from production data to fix one of the category means and then fit the other mean and the category variances. I then compare the found other mean to production data to see if the model fit was successful. But I don't take the production variability into account to inform the modeling process at all. It may be informative to take the production variability as an approximation of the meaningful variance of the categories, or the sum of variances for the categories, as

part of the fitting process. If not to inform the modeling process, then this production variability may be useful in confirming the validity of model fits for the category variances. In situations where neither mean of an individual's categories can be assumed without examining their perception, this might allow us to set both means for the model by eliminating the need to fit the variance parameter, giving us greater flexibility in applying this model. There is currently ongoing work that is using speech corpora to approximate distributions of intended target productions, T values, in order to predict individual's performance on a Same/Different judgment task (N. Feldman, Richter, Falk, & Jansen, 2013). This work will eventually give us a better understanding of how individual variability, production values, behavioral perception data, and the modeling process are inter-related.

Standalone vs Coarticulated Phonemes

The vowel simulations in the L1 part of the dissertation were based on standalone vowels. As discussed in Section 3.4.4, this meant that the primary cue to their identification was their steady state target value for the F1 and F2, which is a static spectral cue. In Section 3.4.5.3, I suggested that static spectral cues, which are used to identify standalone vowels, may be prone to more continuous perception and explain why the standalone vowels were perceived so continuously. Meanwhile, dynamic spectral and static temporal cues, which are shared by fricatives and stop consonants to the exclusion of standalone vowels, may be prone to stronger categorical perception. This is in line with the fact that nasals are also strongly categorically

perceived, and mostly identified by their dynamic spectral cues, with people relying on the static spectral information only in the presence of hard to distinguish formant transitions.

Here I would like to consider how this claim relates to the perception of vowels in coarticulated contexts. We know that coarticulated vowels are identified more readily than standalone ones (Strange, 1989), and while their target values change in different contexts, it is their dynamic behavior that can help listeners identify them accurately. Specifically, if considering /CVS/ syllables, even if we replace the steady state portion of the vowel in the middle with silence, it remains largely identifiable by the remaining dynamic formant transitions (Fox, 1989; Jenkins, Strange, & Trent, 1999). Interestingly, the same formant transitions can be used to identify both the vowel and consonant portions of a syllable (Ohde & German, 2011). This goes against the claim that holds for standalone vowels that their cues to identification do not overlap with those of stop consonants. Beyond just overlapping, it would suggest that vowels in context should be perceived more categorically if dynamic spectral cues really do lead to more categorical behavior. In fact, this is precisely what was found by K. N. Stevens (1966), who showed that one can obtain nearly categorical perception of vowels when considering vowels between consonants pulled out of a rapidly articulated speech stream.

The findings for standalone vowels and this related work on coarticulated vowel perception suggests two avenues for future research. First, having provided further support for the idea that dynamic spectral cues (and temporal ones) may lead to stronger effects of categories, this claim should be investigated with other phoneme

categories, particularly those that in different contexts can lead people to depend more or less on certain cue types. Second, it warrants an application of this model to vowels that are incorporated in /CV/ or /CVC/ syllables, but not just with consonants spliced onto the beginning and end of the vowel, but rather modeled on natural productions with full spectral information, much like was done with the stimuli in the L2 work in this dissertation.

Effects of Other Categories

All the model fits in the L2 section of the dissertation were based on examining the effects of precisely three categories on the perception of experimental stimuli: /i/, /ɨ/, and /ɪ/. However, in both English and Russian there are other vowels that are relatively close to these vowels in F1 and F2 space. Table 5.1 shows that the vowels /e/ and /ɛ/ are located close to the English vowel /ɪ/ in both English and Russian.

If the model accurately captures the source of categorical effects in perception, then in reality all phonemes in the phoneme inventory should be exerting an effect on any speech sound heard. Of course, there are some phonemes that can be easily discarded, like the effect of /ʃ/ on the vowel /u/. Even if we restrict ourselves to just vowels, it is doubtful that the vowel representation of /a/ would have a non-negligible effect on the perception of a stimuli around /i/. And even if the F1/F2 values are similar, these vowels are distinct enough in duration, diphthong structure, and other cues to where they may not be exhibiting a primary categorical

English Vowels					
	Male		Female		
	F1	F2	F1	F2	Source
/i/-English	342	2322	437	2761	J. Hillenbrand et al. (1995)
/ɪ/-English	427	2034	483	2365	J. Hillenbrand et al. (1995)
/e/-English	476	2089	536	2530	J. Hillenbrand et al. (1995)
/ɛ/-English	580	1799	731	2058	J. Hillenbrand et al. (1995)

Russian Vowels			
	F1	F2	Source
/i/-Russian	150	2400	M. Halle and Jones (1959)
/ɨ/-Russian	200	1600	M. Halle and Jones (1959)
/e/-Russian	450	2200	M. Halle and Jones (1959)
/ɛ/-Russian	400	2025	M. Halle and Jones (1959)

Table 5.1: F1 and F2 values for experimental vowels as well as some other vowels close in F1 and F2 valises to those modeled in the L2 part of the dissertation

effect on perception of /ɪ/. However, a fuller simulation would allow for all vowels of a language, or at least a larger set thereof, to play a role. This is not as much of an issue with the L1 simulations since there are not really any other consonants besides /b/ and /p/ that lie on the VOT continuum at this manner and place of articulation. However, certainly along other dimensions like formant transitions and nasalization other consonants are very proximate to these ones. Hence, including more phonemes in the simulations would be a good thing to do all around.

Number of Stimuli in Continuum

A limitation of the specific experiments I carried out for L2 categorical effects is the number of stimuli on any given 1-dimensional continuum. My original goal was to be able to directly model the categories in 2 dimensions. 15 stimuli would have been more than enough data points to do this well, had the model allowed.

However, when it became evident that the primary analysis would be along the three one-dimensional continua of /i/-/ɪ/, /i/-/ɨ/, and /ɪ/-/ɨ/, this ended up only providing 5 points along the continuum, which meant only 4 discrimination scores. All the L1 simulations were based on either 9, 11, or 13 points, partly the reason why I was able to get such good fits and extract accurate category parameters. While I do not believe it kept me from learning important aspects of L2 category change, future work should ensure that any continuum that will be analyzed has more data points to do model fitting on.

The Russian /ɨ/ Vowel

While the goodness and identification data suggest that all endpoints in the two-dimensional continuum of stimuli in the L2 experiments are perceived accurately as good representatives of the relevant vowels, there is some concern over the quality of the stimuli in the /ɨ/ end of the continuum. At least some Russian speakers found that the vowel sounded rather /i/-like, which should not be the case given that /i/ is a separate phoneme in Russian. It was suggested that for Russian speakers, if there was any confusion with this vowel it should be with the English /ɪ/ instead (Kira Gor, personal communication). This may have two sources. First, it might have to do with the fact that the synthesized vowel did not fully resemble the full spectral features of a natively produced one. This is possible since the stimuli in the experiment were based on a native /ɪ/ production and used the formant structure from a single instance of a /ɨ/ produced by a bilingual Russian-English speaker.

Additionally, it might be possible since, according to at least some linguists, the /ɨ/ and /i/ in Russian are allophones of the same underlying phoneme. The status of /ɨ/ as an independent phoneme or allophone of /i/ is a long-standing dispute with the Saint-Petersburg phonology school (Leontev, 1961) holding it to be an independent phoneme while the Moscow school contends that it is an allophone of /i/ (Chew, 2003). If, in fact, the Russian speakers were perceiving the /ɨ/ as more /i/-like than they should, this could have caused some behavior similar to that of native English speakers, who were partially assimilating the /ɨ/ vowel into /i/. This would have lead to results suggesting that advanced learners of Russian were actually closer to native Russians than they really are, providing an overly optimistic view of their category development. An examination of the productions that were collected from both native Russian speakers as well as all learners of Russian could provide a better picture of how far along they really were. This is planned for a future analysis of data collected from these experiments.

5.6 Future Work

L2 Consonant Learning

One area worth pursuing in future work is looking at categorical effects in L2 perception with consonants. Consonants show much stronger effects of categories in native language, so we may be able to see more fine-grained results and changes in parameters when also considering them in newly learned languages. Besides just looking at consonants, future work can tie in what we learn about particular

phonemes in the L1 in terms of the τ values and degree of categorical effects and see what kind of effect this has on rate of learning or rate of approaching native-like processing in the L2. Based on my work so far, I would predict that phonemes with low τ values in L1 processing would lead to more difficulty in establishing new categories in L2 given the greater risk of assimilation due to strong biases toward the category center. If this is the case, then we can start to make predictions that would typically be of the sort made for PAM or SLM in a unified model, much like we collapsed CP and PME in L1 perception in the current work.

Non-Speech Stimuli

Another step in future research would be to apply this model to non-speech processing. This computational model is capable of explaining categorical effects across a range of phonemes and languages, many of which are perceived via drastically different cues. If we do not have to appeal to the neural level of implementation or perhaps even algorithm, and certainly not to the specific measurement, in examining categorical effects, then perhaps it is much more of a domain-general mechanism. Any category that can be represented via a mean and variance, and can be said to have a prior probability should be able to be examined via this model. As a starting point, we can examine various continua that have been shown to have categorical effects, such as color perception (Davidoff et al., 1999), facial expressions (Angeli et al., 2008; Calder et al., 1996), familiar faces (Beale & Keil, 1995), artificial categories of objects (Goldstone et al., 2001), and emotions (Hess et al., 2009; Sauter et al.,

2011). The difficulty would be to identify the proper variable for each continuum. There has already been some related work done in applying similar models to visual dimensions, suggesting that this is a fruitful avenue for investigation. Huttenlocher, Hedges, and Vevea (2000) examined people's reconstruction of fine-grained stimuli within a one-dimensional visual continuum, focusing on the effects of prior knowledge in the form of category distributions. They examined effects on object width, line length, and shades of grey. Like my model which treats speech perception as optimal inference, their Bayesian model treated stimulus reconstruction as a statistical procedure designed to maximize the accuracy of the reproduction. They found that the prior distributions of the categories did in fact affect that bias in people's reconstructions of the objects.

One application of this model that I would find particularly interesting to investigate would be facial recognition. There is a phenomenon whereby people of one race often think that people of another race look very similar, and have trouble identifying specific people or telling two faces apart. These are precisely identification and discrimination tasks. The people we know of various races are the categories in our "face space". Any face we see is pulled toward these categories we know, and if we only have a few categories close to the face we see, those categories will exert the strongest pull, making us perceive something much closer to these categories and warping the perceptual space to make stimuli closer together. If we further stipulate that being less familiar with a certain race introduces perceptual noise, then this would further increase pulls toward category prototypes. This actually falls in line with the idea of establishing prototypes of encountered faces and using them

for future classification that has been supported in research on the cross-race effect (Sporer, 2001). Also, density of exemplars of face clusters has also been investigated for triggering this effect (Valentine & Endo, 1992). I do not believe there is any a-priori reason to say why my model cannot be used to further quantify this type of perception and even examine changes in identification and discrimination that occur with further learning, or exposure to more members of the race in question.

Category Change and L2 Training

Another area of research that naturally comes out of this work is utilizing this model in conjunction with phonetic training for second language learning. Being able to identify what particularly is lacking in the representation of foreign categories, whether an improperly set mean or inappropriate variance, could help design a proper training procedure for specific phonemes. Alternatively, we can use insights from this model to design new training protocols in order to help people with their L2 perception. We can use the model to make quantitative predictions for how speech stimuli will be perceived by listeners based on the categories they possess and the means, variances, and prior probabilities of these categories. If a learner has trouble with a specific new category in a foreign language, this might be due to the fact that their native categories are interfering with setting the "correct" new category. But its possible that the correct new category is not what they need, but rather a slightly different category with a different mean and variance such that acting together they would lead to appropriate perception. Strategies like training listeners on a category

with a mean that overshoots the actual category mean or with a narrower or wider distribution (variance) of the category could be shown to be useful. One type of such training that has been studied is High Variability Phonetic Training (HVPT) (Iverson et al., 2005, inter alia). Under this approach, learners are exposed to repeated forced-choice identification of minimal pairs recorded by multiple speakers and given immediate feedback on their performance. Researchers have shown that this type of training has rapid effects in laboratory settings, particularly in the vowel learning environment (Rato, 2014).

Unfortunately, these types of studies fail to identify exactly what aspect of the perceptual system is adjusted due to the training. It is inconclusive whether the improvement is due to attention paid to the relevant cues as suggested by Iverson et al. (2005), or whether it might have to do with adjustments in finer category parameters like mean and variance due to added exposure. It is reasonable that with exposure to a variety of speakers and contexts, the listeners are better able to estimate the category parameters used for perception. Looking at the task through the lens of my model, it is possible that with increased exposure to variability in a certain category, the listener would end up representing a higher meaningful variance for the category, leading them to pay more attention to acoustic details, and hence to better ability to identify and discriminate the necessary phonemes. In addition, this variability would likely keep them from over-assimilating the stimuli to other categories. Though adjustments to the means are also possible. It would be possible to tell these apart by fitting the model to behavioral performance by speakers before and after such training and seeing which parameters are seen as changing. Future

work would need to be done to explore the relative contribution of the different parameters as well as the relation of the model to other types of phonetic training.

5.7 Conclusion

Assigning perceptual stimuli to categories makes processing more efficient and allows for a discrete hierarchical system such as language to exist. These categories in turn warp the perceptual space by biasing perceived inputs toward category centers. This dissertation examines how effects in both first and second language perception can be brought under the umbrella of a unified Bayesian model. First, I use my model to capture L1 categorical effects for a range of phonemes and establish a τ continuum for gauging the degree of categorical effects for a phoneme or continuum. I then use the same model to track category parameter changes over the course of L2 learning. Together, this work suggests that all categorical effects on phoneme processing in language, or perhaps even domain-generally, may be derivative of a single process that integrates known categories and incoming stimuli via a process of optimal inference of intended targets.

Chapter A: Derivation of identification functions

This appendix provides a detailed derivation of the identification equations used to fit behavioral identification data. First, we observe that the data collected in the behavioral forced-choice identification experiment is an empirical measure of the model probability $p(c_1|S)$, where we observe the rate at which the listener chooses one of the two categories, say c_1, upon observing a speech stimulus, S. According to Bayes' rule (Equation 3.1) we can write this quantity as:

$$p(c_1|S) = \frac{p(S|c_1)p(c_1)}{p(S|c_1)p(c_1) + p(S|c_2)p(c_2)} \tag{A.1}$$

Applying the distributions from the probabilistic model, we have the following representation:

$$p(c_1|S) = \frac{N(\mu_{c1}, \sigma_{c1}^2 + \sigma_S^2)p(c_1)}{N(\mu_{c1}, \sigma_{c1}^2 + \sigma_S^2)p(c_1) + N(\mu_{c2}, \sigma_{c2}^2 + \sigma_S^2)p(c_2)} \tag{A.2}$$

We make the assumption that the two categories in the forced choice identification trial have equal prior probabilities so we replace all prior probabilities with 0.5:

$$p(c_1|S) = \frac{N(\mu_{c1}, \sigma_{c1}^2 + \sigma_S^2)(0.5)}{(0.5)N(\mu_{c1}, \sigma_{c1}^2 + \sigma_S^2) + (0.5)N(\mu_{c2}, \sigma_{c2}^2 + \sigma_S^2)} \tag{A.3}$$

At this point we can divide through by $N(\mu_{c1}, \sigma_{c1}^2 + \sigma_S^2)(0.5)$, giving us:

$$p(c_1|S) = \frac{1}{1 + \frac{N(\mu_{c2}, \sigma_{c2}^2 + \sigma_S^2)}{N(\mu_{c1}, \sigma_{c1}^2 + \sigma_S^2)}} \tag{A.4}$$

If we focus on the term in the denominator, $\frac{N(\mu_{c2}, \sigma_{c2}^2 + \sigma_S^2)}{N(\mu_{c1}, \sigma_{c1}^2 + \sigma_S^2)}$, and replace the normal distributions with the full form of a Gaussian distribution, $N(\mu, \sigma^2) = \sqrt{\frac{1}{2\pi\sigma^2}} exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$, we get (just for this term):

$$\frac{N(\mu_{c2}, \sigma_{c2}^2 + \sigma_S^2)}{N(\mu_{c1}, \sigma_{c1}^2 + \sigma_S^2)} = \frac{\sqrt{\frac{1}{2\pi(\sigma_{c2}^2 + \sigma_S^2)}} exp\left(-\frac{(S-\mu_{c2})^2}{2(\sigma_{c2}^2 + \sigma_S^2)}\right)}{\sqrt{\frac{1}{2\pi(\sigma_{c1}^2 + \sigma_S^2)}} exp\left(-\frac{(S-\mu_{c1})^2}{2(\sigma_{c1}^2 + \sigma_S^2)}\right)} \tag{A.5}$$

Simplifying the square root term, applying the division rule for exponents we get:

$$\sqrt{\frac{(\sigma_{c1}^2 + \sigma_S^2)}{(\sigma_{c2}^2 + \sigma_S^2)}} exp\left(-\frac{(S-\mu_{c2})^2}{2(\sigma_{c2}^2 + \sigma_S^2)} + \frac{(S-\mu_{c1})^2}{2(\sigma_{c1}^2 + \sigma_S^2)}\right) \tag{A.6}$$

Expanding the squares and simplifying we get:

$$\sqrt{\frac{(\sigma_{c1}^2 + \sigma_S^2)}{(\sigma_{c2}^2 + \sigma_S^2)}} exp\left(-\frac{S^2 - 2S\mu_{c2} + \mu_{c2}^2}{2(\sigma_{c2}^2 + \sigma_S^2)} + \frac{S^2 - 2S\mu_{c1} + \mu_{c1}^2}{2(\sigma_{c1}^2 + \sigma_S^2)}\right) \tag{A.7}$$

Converting to a common denominator, simplifying, and plugging back into the original equation gives us:

$$p(c_1|S) = \frac{1}{1 + \sqrt{\frac{\sigma_1^2}{\sigma_2^2}} \times exp\frac{(\sigma_2^2-\sigma_1^2)S^2+2(\mu_{c_2}\sigma_1^2-\mu_{c_1}\sigma_2^2)S+(\mu_{c_1}^2\sigma_2^2-\mu_{c_2}^2\sigma_1^2)}{2\sigma_1^2\sigma_2^2}} \tag{A.8}$$

where $\sigma_1^2 = \sigma_{c_1}^2 + \sigma_S^2$ and $\sigma_2^2 = \sigma_{c_2}^2 + \sigma_S^2$.

Fitting this equation to the behavioral data from our experiments, along with the value for one of the two category means set before the fitting procedure, we are able to derive optimal values for the following four parameters: μ_{c_1}, μ_{c_2}, $\sigma_1^2 = \sigma_{c_1}^2 + \sigma_S^2$, and $\sigma_2^2 = \sigma_{c_2}^2 + \sigma_S^2$.

Chapter B: Derivation of discrimination functions

This appendix lists the detailed derivation of the expected value of the posterior of the target production given the speech sound, $E[T|S]$. By definition of expected value, this is equal to:

$$E[T|S] = \int Tp(T|S)dT \tag{B.1}$$

In our paradigm, the intended target production, T, is potentially derived from a number of categories, particularly in our case, two of them, c_1 and c_2. Hence, we need to marginalize the posterior $p(T|S)$ over the possible categories as $p(T|S) = \sum_c p(T|S,c)p(c|S)$, deriving:

$$E[T|S] = \int T \sum_c p(T|S,c)p(c|S)dT \tag{B.2}$$

Since sums and integrals are order-independent, and since the term $p(c|S)$ does not depend on T, we can rewrite the expected values as:

$$E[T|S] = \sum_c p(c|S) \int Tp(T|S,c)dT \tag{B.3}$$

We already know the term outside of the integral, $p(c|S)$, from the derivation

in Appendix A (Eqs. A.1-A.8). We need to now figure out a way to calculate the inside term, $\int Tp(T|S,c)dT$

Again by definition of expected value, we know that $\int Tp(T|S,c)dT = E[T|S,c]$, so we can rewrite the equation as:

$$E[T|S] = \sum_c p(c|S)E[T|S,c] \tag{B.4}$$

Hence, we can avoid computing the integral if we have another way of calculating the expected value of the intended target production for a specific perceived category. We can do this by considering the distribution $p(T|S,c)$.

Applying Bayes rule, we get the following:

$$p(T|S,c) = \frac{p(S|T)p(T|c)}{\int_T p(S|T)p(T|c)} \tag{B.5}$$

If we rewrite using the actual probability distributions from the model, and remove the normalizing term in the denominator, since we are interested in relative values and not absolutes, we get the following proportional value for $p(T|S,c)$:

$$p(T|S,c) \propto N(T,\sigma_S^2)N(\mu_c,\sigma_c^2) \tag{B.6}$$

We can replace the normal distribution with the equation for the Gaussian, $N(\mu,\sigma^2) = \sqrt{\frac{1}{2\pi\sigma^2}}exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$, in order to get the following representation for $p(T|S,c)$:

$$p(T|S,c) \propto \frac{1}{\sqrt{2\pi\sigma_c^2}} \exp\left(-\frac{(T-\mu_c)^2}{2\sigma_c^2}\right) \frac{1}{\sqrt{2\pi\sigma_S^2}} \exp\left(-\frac{(S-T)^2}{2\sigma_S^2}\right) \tag{B.7}$$

Since we are considering the proportional value we can remove constants that do not depend on T, so removing the two square root normalizing terms and combining exponents together we get:

$$p(T|S,c) \propto \exp\left(-\frac{(T-\mu_c)^2}{2\sigma_c^2} - \frac{(S-T)^2}{2\sigma_S^2}\right) \tag{B.8}$$

Expanding the squared terms and separating the components in the exponent we get the form:

$$p(T|S,c) \propto \exp\left(-\frac{T^2}{2\sigma_c^2} + \frac{2T\mu_c}{2\sigma_c^2} - \frac{\mu_c^2}{2\sigma_c^2} - \frac{S^2}{2\sigma_S^2} + \frac{2ST}{2\sigma_S^2} - \frac{T^2}{2\sigma_S^2}\right) \tag{B.9}$$

We can take the two parts of the exponent that do not depend on T and move them into a separate exponent term:

$$p(T|S,c) \propto \exp\left(-\frac{T^2}{2\sigma_c^2} + \frac{2T\mu_c}{2\sigma_c^2} + \frac{2ST}{2\sigma_S^2} - \frac{T^2}{2\sigma_S^2}\right) \exp\left(-\frac{\mu_c^2}{2\sigma_c^2} - \frac{S^2}{2\sigma_S^2}\right) \tag{B.10}$$

Since this separate term is now just a scalar that does not depend on T, we can remove it entirely, while preserving proportionality, and get the following form:

$$p(T|S,c) \propto \exp\left(-\frac{T^2}{2\sigma_c^2} + \frac{2T\mu_c}{2\sigma_c^2} + \frac{2ST}{2\sigma_S^2} - \frac{T^2}{2\sigma_S^2}\right) \tag{B.11}$$

With a common denominator we get:

$$p(T|S,c) \propto \exp\left(\frac{-\sigma_S^2 T^2 + 2\sigma_S^2 T\mu_c + 2\sigma_c^2 ST - T^2\sigma_c^2}{2\sigma_c^2\sigma_S^2}\right) \tag{B.12}$$

We would like to get this into a form that looks like a Gaussian, so we will need to complete the square. In order to see how to do this we want to group the T^2 and the T terms and see what we need to complete the square. Grouping the terms we get:

$$p(T|S,c) \propto \exp\left(-\frac{\sigma_c^2 + \sigma_S^2}{2(\sigma_c^2\sigma_S^2)}T^2 + \frac{2(\sigma_c^2 S + \sigma_S^2\mu_c)}{2(\sigma_c^2\sigma_S^2)}T\right) \tag{B.13}$$

Now we can isolate the T^2 term by diving through by $\sigma_c^2 + \sigma_S^2$ to get:

$$p(T|S,c) \propto \exp\left(-\frac{T^2 - 2\frac{\sigma_c^2 S + \sigma_S^2\mu_c}{\sigma_c^2+\sigma_S^2}T}{2\frac{\sigma_c^2\sigma_S^2}{\sigma_c^2+\sigma_S^2}}\right) \tag{B.14}$$

We can now complete the square in the numerator by multiplying the equation by a scalar not dependent on T. If we multiply the proportion by $\exp\left(-\frac{\frac{(\sigma_c^2 S+\sigma_S^2\mu_c)^2}{(\sigma_c^2+\sigma_S^2)^2}}{2\frac{\sigma_c^2\sigma_S^2}{\sigma_c^2+\sigma_S^2}}\right)$ and move the exponent term inside we get the form:

$$p(T|S,c) \propto \exp\left(-\frac{T^2 - 2\frac{\sigma_c^2 S + \sigma_S^2\mu_c}{\sigma_c^2+\sigma_S^2}T + \frac{(\sigma_c^2 S+\sigma_S^2\mu_c)^2}{(\sigma_c^2+\sigma_S^2)^2}}{2\frac{\sigma_c^2\sigma_S^2}{\sigma_c^2+\sigma_S^2}}\right) \tag{B.15}$$

This can be rewritten as the complete square:

$$p(T|S,c) \propto \exp\left(-\frac{(T-\frac{\sigma_c^2 S+\sigma_S^2\mu_c}{\sigma_c^2+\sigma_S^2})^2}{2\frac{\sigma_c^2\sigma_S^2}{\sigma_c^2+\sigma_S^2}}\right) \tag{B.16}$$

This now looks precisely like the following normal distribution:

$$p(T|S,c) = N\left(\frac{\sigma_c^2 S+\sigma_S^2\mu_c}{\sigma_c^2+\sigma_S^2}, \frac{\sigma_c^2\sigma_S^2}{\sigma_c^2+\sigma_S^2}\right) \tag{B.17}$$

where the mean is $\frac{\sigma_c^2 S+\sigma_S^2\mu_c}{\sigma_c^2+\sigma_S^2}$ and the variance is $\frac{\sigma_c^2\sigma_S^2}{\sigma_c^2+\sigma_S^2}$.

The expected value $E[T|S,c]$ is precisely the mean of the distribution $p(T|S,c)$. From this normal distribution we have the mean so we have found the expected value:

$$E[T|S,c] = \frac{\sigma_c^2 S+\sigma_S^2\mu_c}{\sigma_c^2+\sigma_S^2} \tag{B.18}$$

We can plug this back into equation B.4 to get:

$$E[T|S] = \sum_c p(c|S)\frac{\sigma_c^2 S+\sigma_S^2\mu_c}{\sigma_c^2+\sigma_S^2} \tag{B.19}$$

For the case of two categories we consider throughout this paper this has the following form:

$$E[T|S] = p(c_1|S)\frac{\sigma_{c1}^2 S+\sigma_S^2\mu_{c1}}{\sigma_{c1}^2+\sigma_S^2} + p(c_2|S)\frac{\sigma_{c2}^2 S+\sigma_S^2\mu_{c2}}{\sigma_{c2}^2+\sigma_S^2} \tag{B.20}$$

The value that is fit from this equation is σ_S^2. Since we already have the sums $\sigma_S^2+\sigma_{c_1}^2$ and $\sigma_S^2+\sigma_{c_2}^2$ from the identification part of the simulations, we can subtract the σ_S^2 term in order to get the individual category variances σ_{c1}^2 and σ_{c2}^2. This gives

us the final parameters needed to derive the fit of the model to the behavioral data for any set of identification and discrimination experiments.

Chapter C: MATLAB fitting procedure for basic simulations

In lieu of printed code in this appendix, all code used in the course of the simulations is available by download from:

`https://dl.dropboxusercontent.com/u/29402237/KronrodDissFiles.zip`

If the code is not available at any point in the future, you can get an electronic copy by emailing me at yakovkronrod@gmail.com or finding me at whatever institution I happen to be at that year. There is only one Yakov Kronrod. You should not have any trouble tracking me down.

Chapter D: Derivation of multidimensional model: Identification

This appendix provides a detailed derivation of the identification equations used to fit behavioral identification data for the multidimensional model. First, we observe that the data collected in the behavioral forced-choice identification experiment is an empirical measure of the model probability $p(c_1|S)$, where we observe the rate at which the listener chooses one of the k categories, say c_1, upon observing a speech stimulus, S. According to Bayes' rule (Equation 3.1) we can write this quantity as:

$$p(c_1|S) = \frac{p(S|c_1)p(c_1)}{\sum_c p(S|c)p(c)} \tag{D.1}$$

In order to get a form for $p(S|c_1)$ we need to consider the distributions from the multidimensional model. In the model, $T \sim N(\mu_c, \Sigma_c)$ and $S \sim N(T, \Sigma_S)$. If we want $p(S|c)$ we can integrate over all possible values of T and describe the distribution as $S|c \sim N(\mu_c, \Sigma_c + \Sigma_S)$. We can now apply the distributions from the probabilistic model and arrive at the following representation:

$$p(c_1|S) = \frac{N(\mu_{c1}, \Sigma_{c1} + \Sigma_S)p(c1)}{\sum_c N(\mu_c, \Sigma_c + \Sigma_S)p(c)} \tag{D.2}$$

If we rewrite $\sum_c N(\mu_c, \Sigma_c + \Sigma_S)p(c) = N(\mu_{c1}, \Sigma_{c1}+\Sigma_S)p(c1)+\sum_{c\neq c1} N(\mu_c, \Sigma_c + \Sigma_S)p(c)$, we get:

$$p(c_1|S) = \frac{N(\mu_{c1}, \Sigma_{c1} + \Sigma_S)p(c1)}{N(\mu_{c1}, \Sigma_{c1} + \Sigma_S)p(c1) + \sum_{c\neq c1} N(\mu_c, \Sigma_c + \Sigma_S)p(c)} \tag{D.3}$$

At this point we can divide through by $N(\mu_{c1}, \Sigma_{c1} + \Sigma_S)p(c_1)$, giving us:

$$p(c_1|S) = \frac{1}{1 + \frac{\sum_{c\neq c1} N(\mu_c, \Sigma_c+\Sigma_S)p(c)}{N(\mu_{c1}, \Sigma_{c1}+\Sigma_S)p(c_1)}} \tag{D.4}$$

We can bring in the $N(\mu_{c1}, \Sigma_{c1} + \Sigma_S)p(c_1)$ term inside the summation, giving us:

$$p(c_1|S) = \frac{1}{1 + \sum_{c\neq c1} \frac{N(\mu_c, \Sigma_c+\Sigma_S)p(c)}{N(\mu_{c1}, \Sigma_{c1}+\Sigma_S)p(c_1)}} \tag{D.5}$$

If we focus on the term in the denominator, $\frac{N(\mu_c, \Sigma_c+\Sigma_S)}{N(\mu_{c1}, \Sigma_{c1}+\Sigma_S)}$, and replace the normal distributions with the full form of a multidimensional Gaussian distribution, $N(\mu, \Sigma) = \frac{1}{(2\pi)^{\frac{k}{2}}|\Sigma|^{\frac{1}{2}}} exp\left[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right]$, we get (just for this term):

$$\frac{N(\mu_c, \Sigma_c + \Sigma_S)}{N(\mu_{c1}, \Sigma_{c1} + \Sigma_S)} = \frac{\frac{1}{(2\pi)^{\frac{k}{2}}|\Sigma_c+\Sigma_S|^{\frac{1}{2}}} exp\left[-\frac{1}{2}(S-\mu_c)^T(\Sigma_c+\Sigma_S)^{-1}(S-\mu_c)\right]]}{\frac{1}{(2\pi)^{\frac{k}{2}}|\Sigma_{c1}+\Sigma_S|^{\frac{1}{2}}} exp\left[-\frac{1}{2}(S-\mu_{c1})^T(\Sigma_{c1}+\Sigma_S)^{-1}(S-\mu_{c1})\right]]} \tag{D.6}$$

We can cancel the 2π terms, combine the determinants, and apply the division rule for exponents, giving us:

$$\frac{N(\mu_c, \Sigma_c + \Sigma_S)}{N(\mu_{c1}, \Sigma_{c1} + \Sigma_S)} = \sqrt{\frac{|\Sigma_{c1} + \Sigma_S|}{|\Sigma_c + \Sigma_S|}} exp\left[-\frac{1}{2}(S-\mu_c)^T(\Sigma_c+\Sigma_S)^{-1}(S-\mu_c) + \frac{1}{2}(S-\mu_{c1})^T(\Sigma_{c1}+\Sigma_S)^{-1}(S-\mu_{c1})\right] \tag{D.7}$$

Plugging this back into the full equation for $p(c_1|S)$, we get:

$$p(c_1|S) = \frac{1}{1 + \sum_{c \neq c1} \frac{p(c)}{p(c_1)} \sqrt{\frac{|\Sigma_{c1}+\Sigma_S|}{|\Sigma_c+\Sigma_S|}} exp\left[EXT_{c\_c1}\right]}$$

$$EXP_{c\_c1} = \left[-\frac{1}{2}(S-\mu_c)^T(\Sigma_c+\Sigma_S)^{-1}(S-\mu_c) + \frac{1}{2}(S-\mu_{c1})^T(\Sigma_{c1}+\Sigma_S)^{-1}(S-\mu_{c1})\right] \tag{D.8}$$

For most simulations, we can make the assumption that all categories are equally likely in the forced choice identification trial and therefore have equal prior probabilities. We lose the $\frac{p(c)}{p(c_1)}$ term and are left with:

$$p(c_1|S) = \frac{1}{1 + \sum_{c \neq c1} \sqrt{\frac{|\Sigma_{c1}+\Sigma_S|}{|\Sigma_c+\Sigma_S|}} exp\left[\frac{(S-\mu_{c1})^T(\Sigma_{c1}+\Sigma_S)^{-1}(S-\mu_{c1}) - (S-\mu_c)^T(\Sigma_c+\Sigma_S)^{-1}(S-\mu_c)}{2}\right]} \tag{D.9}$$

Or, in slightly more elegant terms:

$$p(c_1|S) = \frac{1}{1 + \sum_{c \neq c1} \sqrt{\frac{|\Sigma_{c1\_S}|}{|\Sigma_{c\_S}|}} exp\left[\frac{(S-\mu_{c1})^T \Sigma_{c1\_S}^{-1}(S-\mu_{c1}) - (S-\mu_c)^T \Sigma_{c\_S}^{-1}(S-\mu_c)}{2}\right]} \tag{D.10}$$

where $\Sigma_{c\_S} = \Sigma_c + \Sigma_S$ and $\Sigma_{c1\_S} = \Sigma_{c1} + \Sigma_S$

Fitting this equation to the behavioral data from experiments with multiple dimensions we are able to derive optimal values for the means and covariance matrices representing the sum of the noise and meaningful category covariance, $\Sigma_c + \Sigma_S$, of an arbitrary number of categories. However, as detailed in Chapter 4, multiple parameters need to be set in order to converge on a single optimal solution.

Chapter E: Derivation of multidimensional model: Discrimination

This appendix provides a detailed derivation of the discrimination equations used to fit behavioral discrimination data for the multidimensional model. Specifically, I derive the expected value of the posterior of the target production given the speech sound, $E[T|S]$. By definition of expected value, this is equal to:

$$E[T|S] = \int_k T p(T|S) dT \tag{E.1}$$

[where $\int_k$ denotes an integral over k dimensions]

In our paradigm, the intended target production, T, is potentially derived from a number of categories. Hence, we need to marginalize the posterior $p(T|S)$ over the possible categories as $p(T|S) = \sum_c p(T|S,c)p(c|S)$, deriving:

$$E[T|S] = \int_k T \sum_c p(T|S,c)p(c|S) dT \tag{E.2}$$

Since sums and integrals are order-independent, and since the term $p(c|S)$ does not depend on T, we can rewrite the expected values as:

$$E[T|S] = \sum_c p(c|S) \int_k T p(T|S,c) dT \tag{E.3}$$

We already know the term outside of the integral, $p(c|S)$, from the derivation in Appendix D. We need to now figure out a way to calculate the inside term, $\int_k Tp(T|S,c)dT$

Again by definition of expected value, we know that $\int_k Tp(T|S,c)dT = E[T|S,c]$, so we can rewrite the equation as:

$$E[T|S] = \sum_c p(c|S)E[T|S,c] \tag{E.4}$$

Hence, we can avoid computing the integral if we have another way of calculating the expected value of the intended target production for a specific perceived category. We can do this by considering the distribution $p(T|S,c)$ and seeing if we can get it into a single Normal form from which we can easily extract the expected value of the distribution.

Applying Bayes rule, we get the following:

$$p(T|S,c) = \frac{p(S|T)p(T|c)}{\int_T p(S|T)p(T|c)} \tag{E.5}$$

If we rewrite using the actual probability distributions from the model, and remove the normalizing term in the denominator, since we are interested in relative values and not absolutes, we get the following proportional value for $p(T|S,c)$:

$$p(T|S,c) \propto N(T,\Sigma_S)N(\mu_c,\Sigma_c) \tag{E.6}$$

We can replace the normal distribution with the equation for the multidimen-

sional Gaussian:

$$N(\mu, \Sigma) = \frac{1}{(2\pi)^{\frac{k}{2}}|\Sigma|^{\frac{1}{2}}} exp\left[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right] \tag{E.7}$$

in order to get the following representation for $p(T|S,c)$:

$$\begin{aligned} p(T|S,c) \propto & \frac{1}{(2\pi)^{\frac{k}{2}}|\Sigma_S|^{\frac{1}{2}}} exp\left[-\frac{1}{2}(S-T)^T\Sigma_S^{-1}(S-T)\right] \times \\ & \frac{1}{(2\pi)^{\frac{k}{2}}|\Sigma_c|^{\frac{1}{2}}} exp\left[-\frac{1}{2}(T-\mu_c)^T\Sigma_c^{-1}(T-\mu_c)\right] \end{aligned} \tag{E.8}$$

Since we are considering the proportional value we can remove constants that do not depend on T, giving us:

$$p(T|S,c) \propto exp\left[-\frac{1}{2}(S-T)^T\Sigma_S^{-1}(S-T)\right] \times exp\left[-\frac{1}{2}(T-\mu_c)^T\Sigma_c^{-1}(T-\mu_c)\right] \tag{E.9}$$

We can now combine the exponents together to get the form:

$$p(T|S,c) \propto exp\left[-\frac{1}{2}(S-T)^T\Sigma_S^{-1}(S-T) - \frac{1}{2}(T-\mu_c)^T\Sigma_c^{-1}(T-\mu_c)\right] \tag{E.10}$$

We can expand the two parts of the matrix multiplications inside the exponent as:

$$
\begin{aligned}
(S-T)^T\Sigma_S^{-1}(S-T) &= (S^T-T^T)\Sigma_S^{-1}(S-T) \\
&= (S^T\Sigma_S^{-1}-T^T\Sigma_S^{-1})(S-T) \\
&= S^T\Sigma_S^{-1}S - S^T\Sigma_S^{-1}T - T^T\Sigma_S^{-1}S + T^T\Sigma_S^{-1}T \\
(T-\mu_c)^T\Sigma_c^{-1}(T-\mu_c) &= (T^T-\mu_c^T)\Sigma_c^{-1}(T-\mu_c) \\
&= (T^T\Sigma_c^{-1}-\mu_c^T\Sigma_c^{-1})(T-\mu_c) \\
&= T^T\Sigma_c^{-1}T - T^T\Sigma_c^{-1}\mu_c - \mu_c^T\Sigma_c^{-1}T + \mu_c^T\Sigma_c^{-1}\mu_c
\end{aligned}
$$

Plugging these into the exponent and pulling out the $-\frac{1}{2}$ term we get:

$$
\begin{aligned}
p(T|S,c) \propto exp\Big[-\frac{1}{2}\big(&S^T\Sigma_S^{-1}S - S^T\Sigma_S^{-1}T - T^T\Sigma_S^{-1}S + T^T\Sigma_S^{-1}T \\
&+T^T\Sigma_c^{-1}T - T^T\Sigma_c^{-1}\mu_c - \mu_c^T\Sigma_c^{-1}T + \mu_c^T\Sigma_c^{-1}\mu_c\big)\Big]
\end{aligned}
\tag{E.11}
$$

We can take the two terms that don't depend on T, $S^T\Sigma_S^{-1}S$ and $\mu_c^T\Sigma_c^{-1}\mu_c$, and move them into a separate exponent term:

$$
\begin{aligned}
p(T|S,c) \propto exp\Big[-\frac{1}{2}\big(&-S^T\Sigma_S^{-1}T - T^T\Sigma_S^{-1}S + T^T\Sigma_S^{-1}T \\
&+T^T\Sigma_c^{-1}T - T^T\Sigma_c^{-1}\mu_c - \mu_c^T\Sigma_c^{-1}T\big)\Big] \times \\
&exp\Big[-\frac{1}{2}\big(S^T\Sigma_S^{-1}S + \mu_c^T\Sigma_c^{-1}\mu_c\big)\Big]
\end{aligned}
\tag{E.12}
$$

Since this separate term is now just a scalar that does not depend on T, we

can remove it entirely, while preserving proportionality, and get the following term:

$$p(T|S,c) \propto exp\left[-\frac{1}{2}\left(-S^T\Sigma_S^{-1}T - T^T\Sigma_S^{-1}S + T^T\Sigma_S^{-1}T \right.\right. \\ \left.\left. +T^T\Sigma_c^{-1}T - T^T\Sigma_c^{-1}\mu_c - \mu_c^T\Sigma_c^{-1}T\right)\right] \tag{E.13}$$

We want to get $p(T|S,c)$ into a form that resembles the multidimensional gaussian presented in Equation E.7. To do this we need to get the equation into a form where we can split it up into $(T - something)^T$ multiplied by the inverse of a variance term and then multiplied by $T - something$). To start we will group terms that have the same T terms, giving us:

$$p(T|S,c) \propto exp\left[-\frac{1}{2}\left((T^T\Sigma_S^{-1}T + T^T\Sigma_c^{-1}T)\right.\right. \\ \left.\left. -(S^T\Sigma_S^{-1}T + \mu_c^T\Sigma_c^{-1}T) - (T^T\Sigma_S^{-1}S + T^T\Sigma_c^{-1}\mu_c)\right)\right] \tag{E.14}$$

We can rewrite this as:

$$p(T|S,c) \propto exp\left[-\frac{1}{2}\left(T^T(\Sigma_S^{-1} + \Sigma_c^{-1})T \right.\right. \\ \left.\left. -(S^T\Sigma_S^{-1} + \mu_c^T\Sigma_c^{-1})T - T^T(\Sigma_S^{-1}S + \Sigma_c^{-1}\mu_c)\right)\right] \tag{E.15}$$

For space consideration, I will consider just the term inside the exponent after the $\frac{1}{2}$ term, continuing my derivation with:

$$\propto T^T(\Sigma_S^{-1} + \Sigma_c^{-1})T - (S^T\Sigma_S^{-1} + \mu_c^T\Sigma_c^{-1})T - T^T(\Sigma_S^{-1}S + \Sigma_c^{-1}\mu_c) \tag{E.16}$$

I want to be able to pull out a $T^T(\Sigma_S^{-1} + \Sigma_c^{-1})$ term so I will add the identity matrix $I_k = (\Sigma_S^{-1} + \Sigma_c^{-1})(\Sigma_S^{-1} + \Sigma_c^{-1})^{-1}$ into the last term and move it to the other T^T term to get:

$$\propto T^T(\Sigma_S^{-1} + \Sigma_c^{-1})T - T^T(\Sigma_S^{-1} + \Sigma_c^{-1})(\Sigma_S^{-1} + \Sigma_c^{-1})^{-1}(\Sigma_S^{-1}S + \Sigma_c^{-1}\mu_c) - (S^T\Sigma_S^{-1} + \mu_c^T\Sigma_c^{-1})T \tag{E.17}$$

We want to split into two binomials, similar to completing a square. Hence we will need to add a constant to make the math work out. We can add a constant that is not dependent on T since we are dealing with a proportion and not an exact solution. The constant will be the multiple of the term preceding T in the last term, $(S^T\Sigma_S^{-1} + \mu_c^T\Sigma_c^{-1})$ and the term following $T^T(\Sigma_S^{-1} + \Sigma_c^{-1})$ in the second term, $(\Sigma_S^{-1} + \Sigma_c^{-1})^{-1}(\Sigma_S^{-1}S + \Sigma_c^{-1}\mu_c)$.

This gives us the full equation:

$$\begin{aligned} &\propto T^T(\Sigma_S^{-1} + \Sigma_c^{-1})T - T^T(\Sigma_S^{-1} + \Sigma_c^{-1})(\Sigma_S^{-1} + \Sigma_c^{-1})^{-1}(\Sigma_S^{-1}S + \Sigma_c^{-1}\mu_c) \\ &-(S^T\Sigma_S^{-1} + \mu_c^T\Sigma_c^{-1})T + (S^T\Sigma_S^{-1} + \mu_c^T\Sigma_c^{-1})(\Sigma_S^{-1} + \Sigma_c^{-1})^{-1}(\Sigma_S^{-1}S + \Sigma_c^{-1}\mu_c) \end{aligned} \tag{E.18}$$

Splitting into the binomial we now get:

$$\begin{aligned} &\propto \left[T^T(\Sigma_S^{-1}+\Sigma_c^{-1}) - (S^T\Sigma_S^{-1}+\mu_c^T\Sigma_c^{-1})\right] \times \\ &\left[T - (\Sigma_S^{-1}+\Sigma_c^{-1})^{-1}(\Sigma_S^{-1}S+\Sigma_c^{-1}\mu_c)\right] \end{aligned} \tag{E.19}$$

I want to be able to isolate the T term in the left binomial just like in the right one, so I need to be able to pull out the $(\Sigma_S^{-1}+\Sigma_c^{-1}$ term from both terms in the binomial. To do this I again add an identity matrix, $I_k = (\Sigma_S^{-1}+\Sigma_c^{-1})^{-1}(\Sigma_S^{-1}+\Sigma_c^{-1})$ to the second term of the binomial, giving us:

$$\begin{aligned} \propto \left[T^T(\Sigma_S^{-1}+\Sigma_c^{-1}) - (S^T\Sigma_S^{-1}+\mu_c^T\Sigma_c^{-1})(\Sigma_S^{-1}+\Sigma_c^{-1})^{-1}(\Sigma_S^{-1}+\Sigma_c^{-1})\right] \times \\ \left[T - (\Sigma_S^{-1}+\Sigma_c^{-1})^{-1}(\Sigma_S^{-1}S+\Sigma_c^{-1}\mu_c)\right] \end{aligned} \tag{E.20}$$

I can now pull out the $(\Sigma_S^{-1}+\Sigma_c^{-1})$ term and get a form starting to look like a multidimensional gaussian:

$$\begin{aligned} &\propto \left[T^T - (S^T\Sigma_S^{-1}+\mu_c^T\Sigma_c^{-1})(\Sigma_S^{-1}+\Sigma_c^{-1})^{-1}\right] \times (\Sigma_S^{-1}+\Sigma_c^{-1}) \times \\ &\left[T - (\Sigma_S^{-1}+\Sigma_c^{-1})^{-1}(\Sigma_S^{-1}S+\Sigma_c^{-1}\mu_c)\right] \end{aligned} \tag{E.21}$$

I now want to pull out the transpose term from the left binomial. Before I can bring the transpose term outside the brackets, I need to pull it out of the first term using the identify $(AB)^T = B^TA$:

$$\propto \left[T^T - (\Sigma_S^{-1}S + \Sigma_c^{-1}\mu_c)^T(\Sigma_S^{-1} + \Sigma_c^{-1})^{-1}\right] \times (\Sigma_S^{-1} + \Sigma_c^{-1})\times \left[T - (\Sigma_S^{-1} + \Sigma_c^{-1})^{-1}(\Sigma_S^{-1}S + \Sigma_c^{-1}\mu_c)\right] \tag{E.22}$$

Applying the same identity to the whole second term of the left binomial gives us:

$$\propto \left[T^T - [(\Sigma_S^{-1} + \Sigma_c^{-1})^{-1}(\Sigma_S^{-1}S + \Sigma_c^{-1}\mu_c)]^T\right] \times (\Sigma_S^{-1} + \Sigma_c^{-1})\times \left[T - (\Sigma_S^{-1} + \Sigma_c^{-1})^{-1}(\Sigma_S^{-1}S + \Sigma_c^{-1}\mu_c)\right] \tag{E.23}$$

I can now pull the transpose term completely out of the first binomial, giving the form:

$$\propto \left[T - (\Sigma_S^{-1} + \Sigma_c^{-1})^{-1}(\Sigma_S^{-1}S + \Sigma_c^{-1}\mu_c)\right]^T \times (\Sigma_S^{-1} + \Sigma_c^{-1})\times \left[T - (\Sigma_S^{-1} + \Sigma_c^{-1})^{-1}(\Sigma_S^{-1}S + \Sigma_c^{-1}\mu_c)\right] \tag{E.24}$$

Now I can plug this whole term back into the full exponent form to get the full equation:

$$p(T|S,c) \propto exp\left[-\frac{1}{2}[T-(\Sigma_S^{-1}+\Sigma_c^{-1})^{-1}(\Sigma_S^{-1}S+\Sigma_c^{-1}\mu_c)]^T \times (\Sigma_S^{-1}+\Sigma_c^{-1})\times \right.$$
$$\left.[T-(\Sigma_S^{-1}+\Sigma_c^{-1})^{-1}(\Sigma_S^{-1}S+\Sigma_c^{-1}\mu_c)]\right] \tag{E.25}$$

In order to get it to look like the gaussian we can add a constant term that does not depend on T, giving us the final form of the posterior:

$$p(T|S,c) \propto \frac{1}{(2\pi)^{\frac{k}{2}}|(\Sigma_S^{-1}+\Sigma_c^{-1})^{-1}|^{\frac{1}{2}}} exp\left[-\frac{1}{2}[T-(\Sigma_S^{-1}+\Sigma_c^{-1})^{-1}(\Sigma_S^{-1}S+\Sigma_c^{-1}\mu_c)]^T \times \right.$$
$$\left.(\Sigma_S^{-1}+\Sigma_c^{-1}) \times [T-(\Sigma_S^{-1}+\Sigma_c^{-1})^{-1}(\Sigma_S^{-1}S+\Sigma_c^{-1}\mu_c)]\right] \tag{E.26}$$

This is now precisely the form of a multidimensional gaussian, $N(\mu,\Sigma) = \frac{1}{(2\pi)^{\frac{k}{2}}|\Sigma|^{\frac{1}{2}}} exp\left[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right]$, with $\mu = (\Sigma_S^{-1}+\Sigma_c^{-1})^{-1}(\Sigma_S^{-1}S+\Sigma_c^{-1}\mu_c)$ and $\Sigma = (\Sigma_S^{-1}+\Sigma_c^{-1})^{-1}$. We can rewrite as:

$$p(T|S,c) \propto N\left(\left[(\Sigma_S^{-1}+\Sigma_c^{-1})^{-1}(\Sigma_S^{-1}S+\Sigma_c^{-1}\mu_c)\right], \left[(\Sigma_S^{-1}+\Sigma_c^{-1})^{-1}\right]\right) \tag{E.27}$$

The expected value $E[T|S,c]$ is precisely the mean of the distribution $p(T|S,c)$. From this normal distribution we have the mean so we have found the expected value:

$$E[T|S,c] = (\Sigma_S^{-1} + \Sigma_c^{-1})^{-1}(\Sigma_S^{-1}S + \Sigma_c^{-1}\mu_c) \tag{E.28}$$

We can plug this back into Equation E.4 to get the expected value of the full posterior:

$$E[T|S] = \sum_c p(c|S)(\Sigma_S^{-1} + \Sigma_c^{-1})^{-1}(\Sigma_S^{-1}S + \Sigma_c^{-1}\mu_c) \tag{E.29}$$

Fitting this equation to the behavioral discrimination data from experiments with multiple dimensions we are able to derive an optimal value for the noise covariance matrix, Σ_S. This in turn gives us the values for all meaningful underlying categorical covariance matrices since all sums $\Sigma_c + \Sigma_S$ are already fit in the identification section of the procedure. The particular fitting procedure and error minimization routine may differ depending on the type of data available (d-prime vs. accuracy data), but this equation gives us the ability to calculate the expected (and hence, inferred) target production value given any number of categories.

Chapter F: Materials for the L2 experiment, full formant grids

The actual final stimuli that were used for the experiments with native English speaking L2 learners of Russian can be downloaded by visiting:

`https://dl.dropboxusercontent.com/u/29402237/KronrodDissFiles.zip`

You can also email me at yakovkronrod@gmail.com if you have trouble retrieving any relevant files. Below are the parameters used in creating the stimuli as well as the pairings of stimuli used in the discrimination experiment.

F.1 Weights for calculating stimuli formant grids

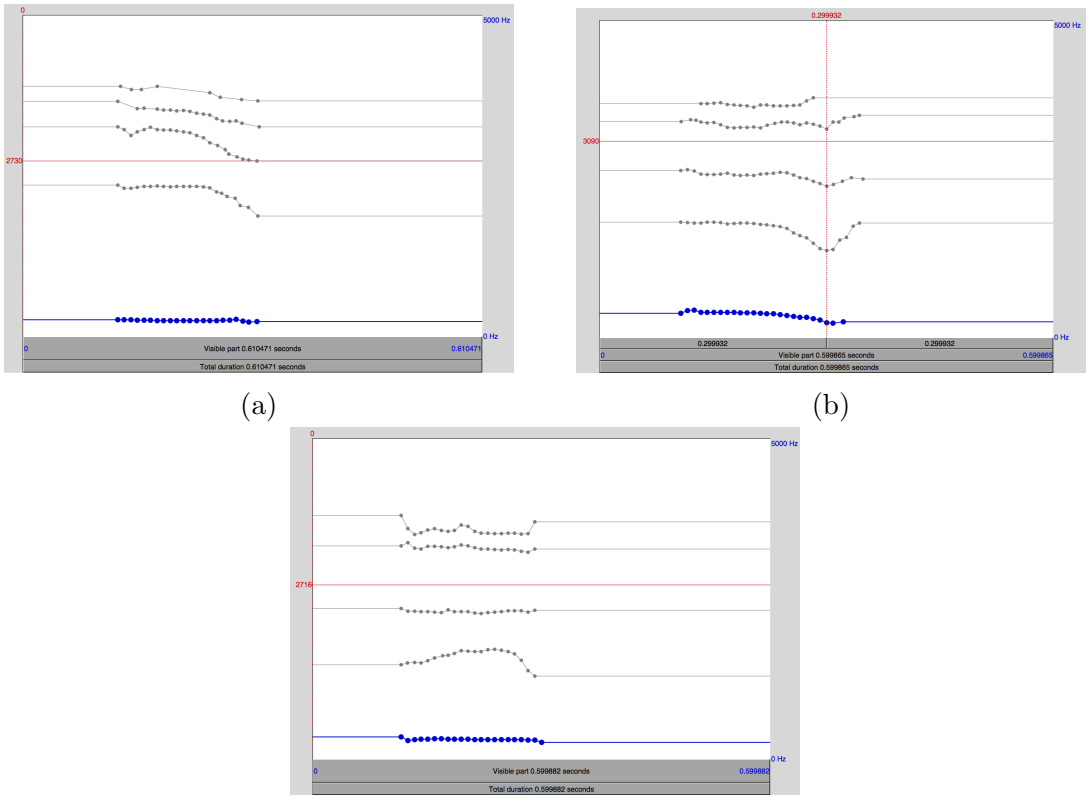
The weights for combining the three corner stimuli into the final stimuli for the L2 experiemnts are provided below in Table F.1.

F.2 Full stimuli formant grids

Figure F.1 contains the formant grids for the three corner stimuli used in linear combination to create all the stimuli for the experiment.

Stimulus	Target F1	Target F2	/i/ weight	/ɪ/ weight	/ɨ/ weight
s11	250	2319	0.985	0.028	-0.013
s13	250	2112	0.719	-0.163	0.444
s15	250	1924	0.477	-0.337	0.86
s17	250	1752	0.256	-0.496	1.24
s19	250	1595	0.054	-0.641	1.586
s22	297	2213	0.738	0.453	-0.191
s24	297	2016	0.485	0.271	0.244
s26	297	1836	0.253	0.105	0.642
s28	297	1672	0.042	-0.046	1.004
s33	347	2112	0.49	0.916	-0.407
s35	347	1924	0.249	0.743	0.009
s37	347	1752	0.027	0.584	0.389
s44	399	2016	0.244	1.406	-0.65
s46	399	1836	0.013	1.24	-0.253
s55	453	1924	-0.001	1.922	-0.921

Table F.1: These weights are used to linearly combine the three naturally produced stimuli together to create all the stimuli along the two-dimensional continuum for the L2 experiments. The target F1 and F2 values are based on the central measure of the vowels in F1/F2 space. By combining the three productions we get fully specified natural sounding diphthong structure throughout the vowel while still maintaining equal spacing as measured by their central stable formants.



(a) (b)

(c)

Figure F.1: Full formant grids for the original stimuli used in the process of stimuli creation for the L2 experiments. The formant grids of these three productions are combined according to the weights calculated for every stimulus in the 2-dimensional continuum. Formant grids correspond to: (a) /i/, (b) /ɪ/, and (c) /ɨ/

F.3 Discrimination Pairs

Table F.2 contains all the discrimination pairs used in the Discrimination experiment for L2 learners of Russian. The condition name contains the reference item and the directions for how to get to the second item, with S=Same, D=Down, R=Right, and L=Left. For clarity, I also include the specific items that belong to each pair.

Condition	item1	item2	Condition	Item1	Item2	Condition	Item1	Item2
Same Pairs			1-Step			2-Step		
11S	11	11	11RR	11	13	11RRRR	11	15
13S	13	13	11DR	11	22	11DRRR	11	24
15S	15	15	13RR	13	15	11DDRR	11	33
17S	17	17	13DL	13	22	13RRRR	13	17
19S	19	19	13DR	13	24	13DRRR	13	26
22S	22	22	13DD	13	33	13DDRR	13	35
24S	24	24	15RR	15	17	15RRRR	15	19
26S	26	26	15DL	15	24	15DLLL	15	22
28S	28	28	15DR	15	26	15DRRR	15	28
33S	33	33	15DD	15	35	15DDLL	15	33
35S	35	35	17RR	17	19	15DDRR	15	37
37S	37	37	17DL	17	26	17DLLL	17	24
44S	44	44	17DR	17	28	17DDLL	17	35
46S	46	46	17DD	17	37	19DLLL	19	26
55S	55	55	19DL	19	28	19DDLL	19	37
			22RR	22	24	22RRRR	22	26
			22DR	22	33	22DRRR	22	35
			24RR	24	26	22DDRR	22	44
			24DL	24	33	24RRRR	24	28
			24DR	24	35	24DRRR	24	37
			24DD	24	44	24DDRR	24	46
			26RR	26	28	26DLLL	26	33
			26DL	26	35	26DDLL	26	44
			26DR	26	37	28DLLL	28	35
			26DD	26	46	28DDLL	28	46
			28DL	28	37	33RRRR	33	37
			33RR	33	35	33DRRR	33	46
			33DR	33	44	33DDRR	33	55
			35RR	35	37	37DLLL	37	44
			35DL	35	44	37DDLL	37	55
			35DR	35	46			
			35DD	35	55			
			37DL	37	46			
			44RR	44	46			
			44DR	44	55			
			46DL	46	55			

Table F.2: All pairs of stimuli used in the discrimination experiment

Chapter G: Full set of findings in the L2 experiment

This appendix contains all the raw results for the three experiments on native English speaking L2 learners of Russian.

G.1 Goodness Distributions

Figures G.5 - **??** show the goodness ratings distributions for all levels of speakers.

G.2 Identification Findings

Figures G.6 - G.10 show the identification findings for all levels of speakers.

G.3 Discrimination Findings

In Tables G.1-**??** I present the raw discrimination results for all pairs tested in the experiment. The scores presented are accuracy scores grouped by level of speakers. By taking the accuracy scores for the same and different pairs, d' scores can in turn be calculated for any set of pairs for further analysis.

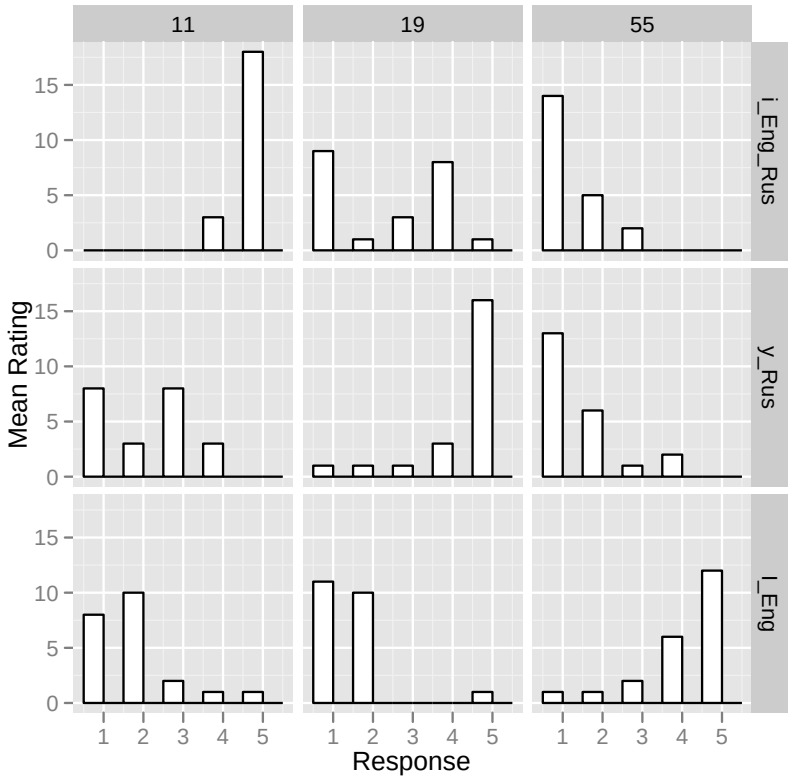


Figure G.1: Distribution of goodness ratings by naive native English speakers. Judgments are on a 1-5 scale for each of the three corner stimuli as representative of the three vowel categories /i/ (i_Eng_Rus), /ɨ/ (y_Rus), and /ɪ/ (I_Eng). Distributions along the diagonal represent ratings of corner stimuli as the vowel they are meant to represent. The other squares in the plot represent judgments of the corner stimuli as the other two vowels

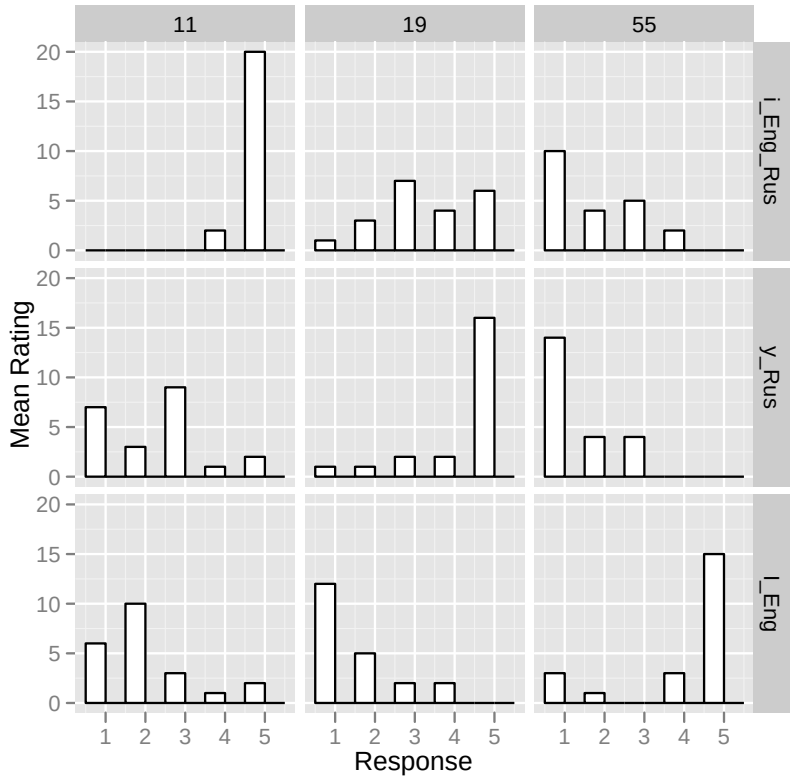


Figure G.2: Distribution of goodness ratings by beginner learners of Russian. Judgments are on a 1-5 scale for each of the three corner stimuli as representative of the three vowel categories /i/ (i_Eng_Rus), /ɨ/ (y_Rus), and /ɪ/ (I_Eng). Distributions along the diagonal represent ratings of corner stimuli as the vowel they are meant to represent. The other squares in the plot represent judgments of the corner stimuli as the other two vowels

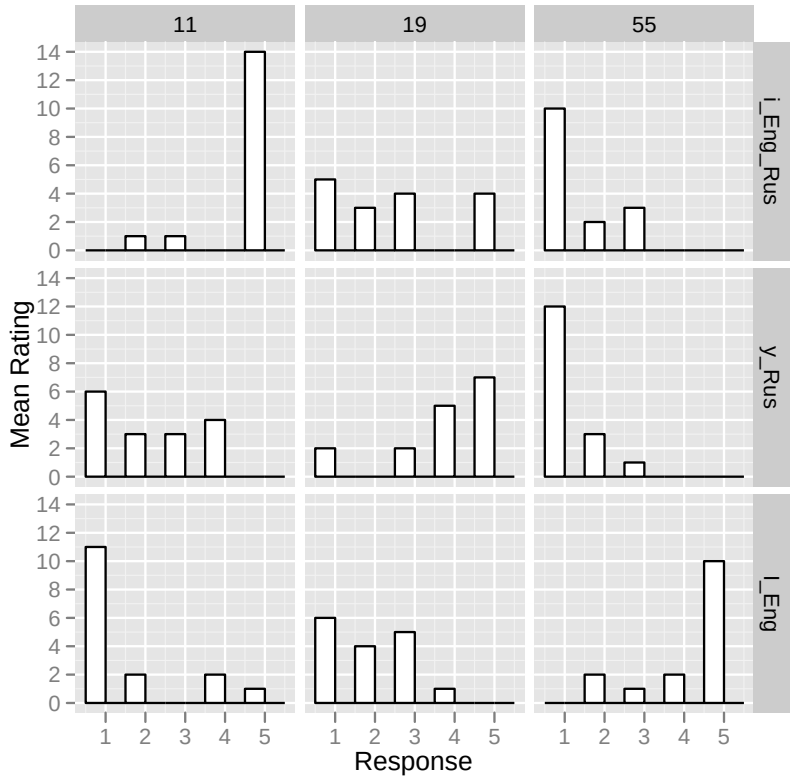


Figure G.3: Distribution of goodness ratings by intermediate learners of Russian. Judgments are on a 1-5 scale for each of the three corner stimuli as representative of the three vowel categories /i/ (i_Eng_Rus), /ɨ/ (y_Rus), and /ɪ/ (I_Eng). Distributions along the diagonal represent ratings of corner stimuli as the vowel they are meant to represent. The other squares in the plot represent judgments of the corner stimuli as the other two vowels

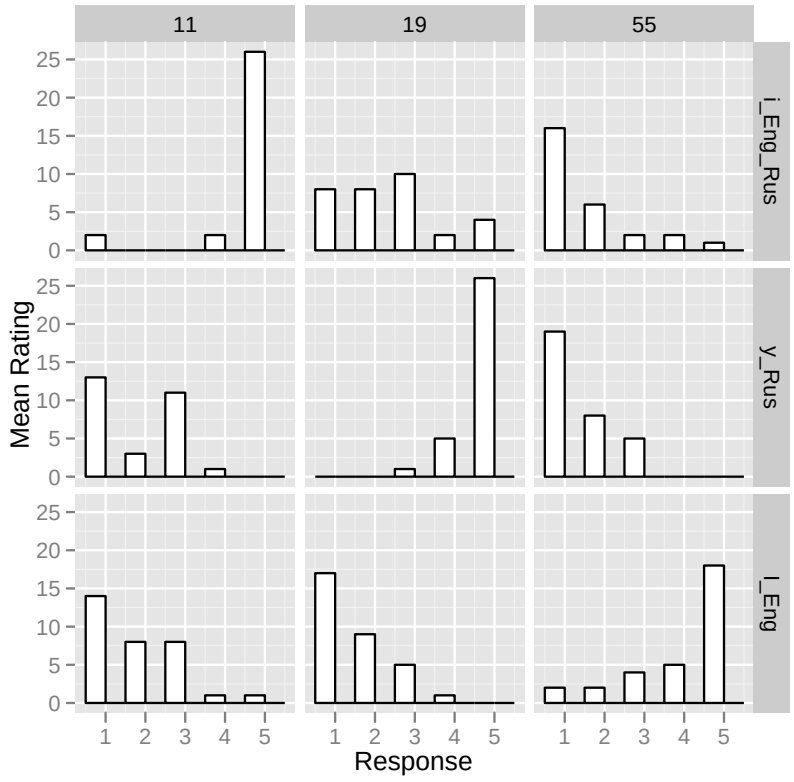


Figure G.4: Distribution of goodness ratings by advanced learners of Russian. Judgments are on a 1-5 scale for each of the three corner stimuli as representative of the three vowel categories /i/ (i_Eng_Rus), /ɨ/ (y_Rus), and /ɪ/ (I_Eng). Distributions along the diagonal represent ratings of corner stimuli as the vowel they are meant to represent. The other squares in the plot represent judgments of the corner stimuli as the other two vowels

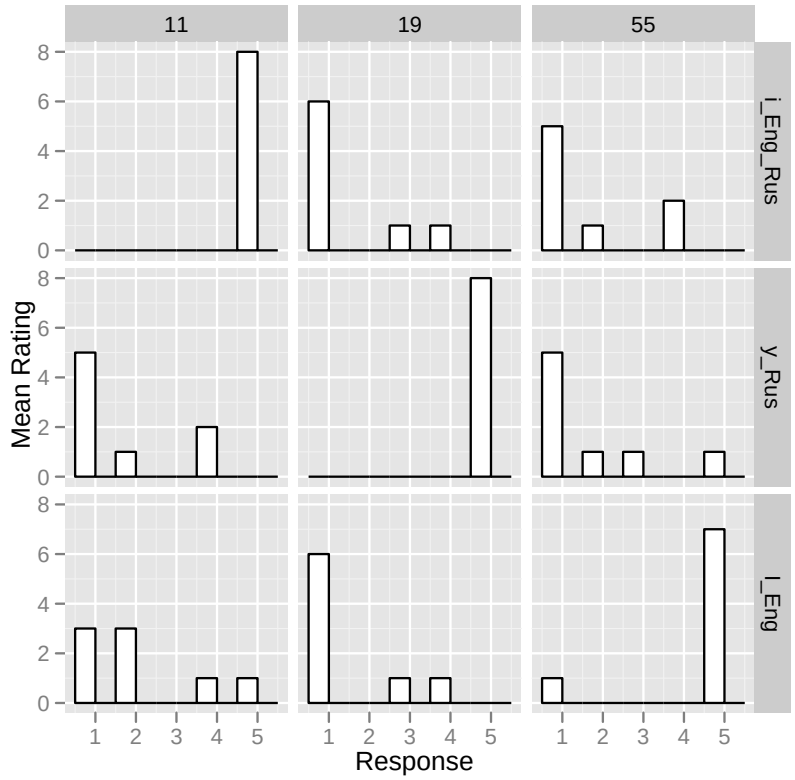


Figure G.5: Distribution of goodness ratings by native Russian speakers. Judgments are on a 1-5 scale for each of the three corner stimuli as representative of the three vowel categories /i/ (i_Eng_Rus), /ɨ/ (y_Rus), and /ɪ/ (I_Eng). Distributions along the diagonal represent ratings of corner stimuli as the vowel they are meant to represent. The other squares in the plot represent judgments of the corner stimuli as the other two vowels

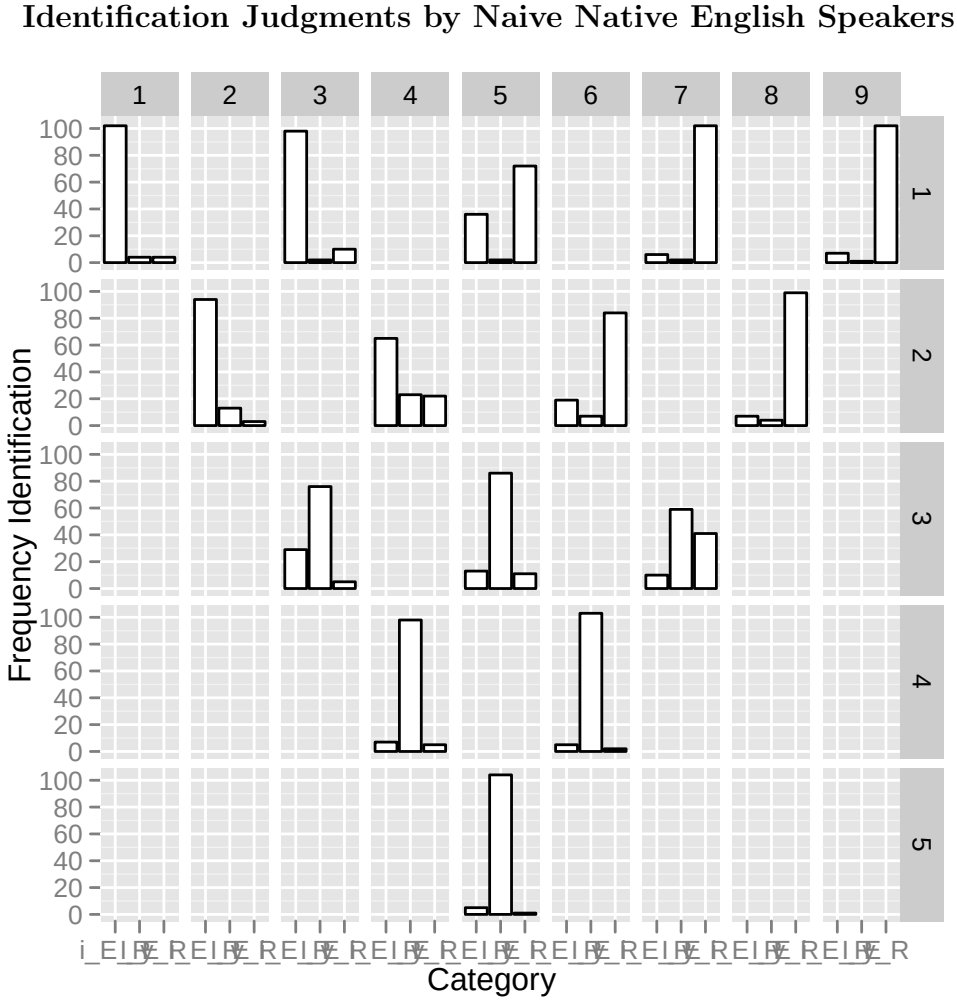


Figure G.6: Distribution of identification judgments by naive native English speakers for all stimuli in the 2D continuum. Each histogram represents, from left to right, classification as /i/, /ɪ/, and /ɨ/.

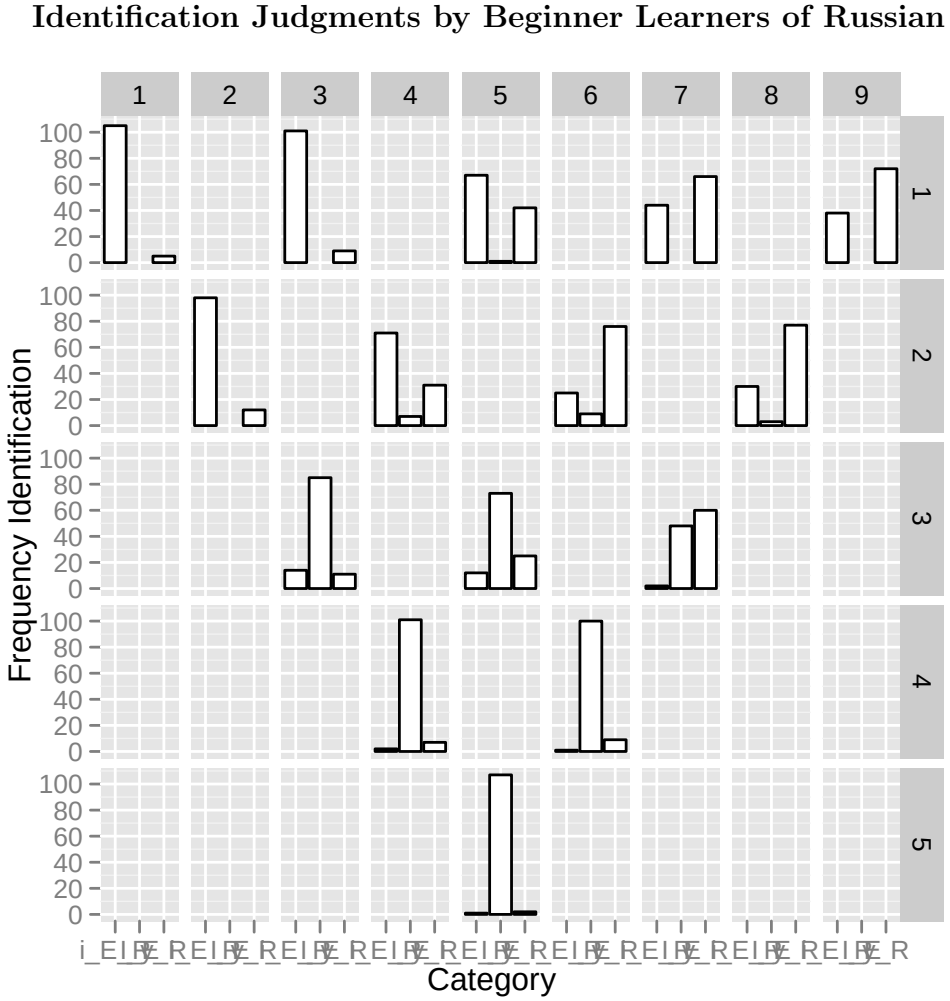


Figure G.7: Distribution of identification judgments by beginner learners of Russian for all stimuli in the 2D continuum. Each histogram represents, from left to right, classification as /i/, /ɪ/, and /ɨ/.

Identification Judgments by Intermediate Learners of Russian

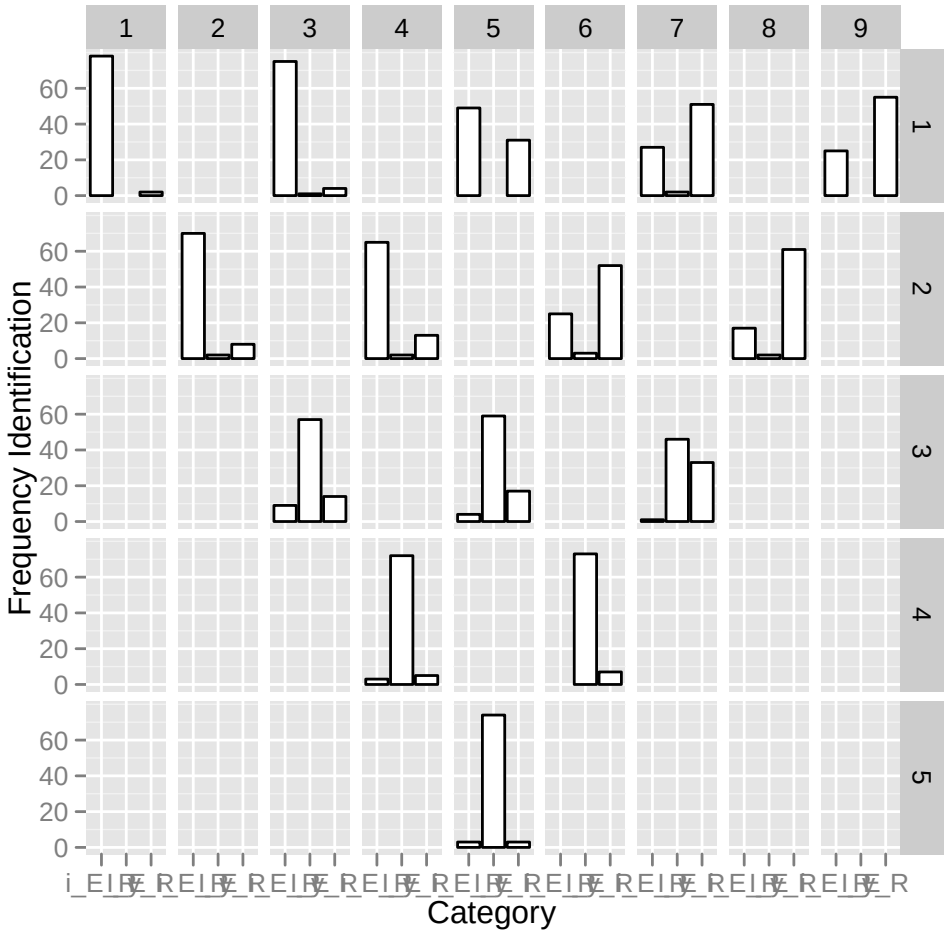


Figure G.8: Distribution of identification judgments by intermediate learners of Russian for all stimuli in the 2D continuum. Each histogram represents, from left to right, classification as /i/, /ɪ/, and /ɨ/.

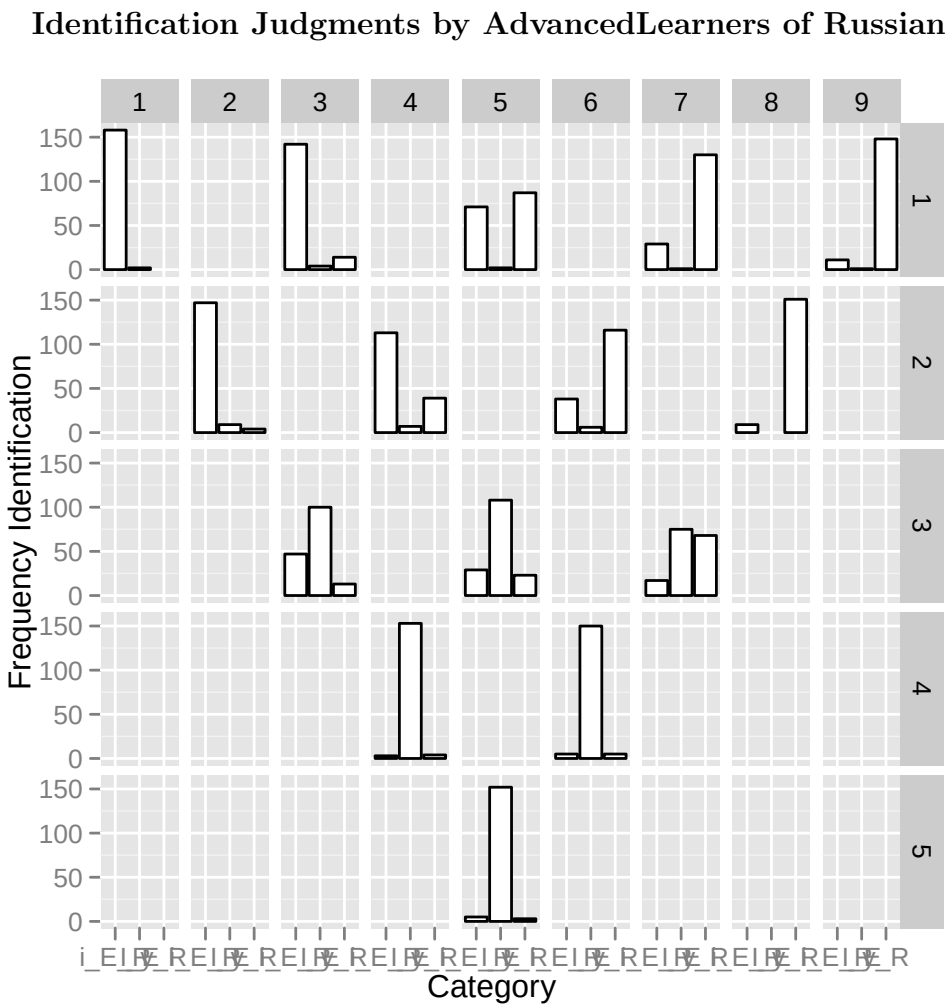


Figure G.9: Distribution of identification judgments by advanced learners of Russian for all stimuli in the 2D continuum. Each histogram represents, from left to right, classification as /i/, /ɪ/, and /ɨ/.

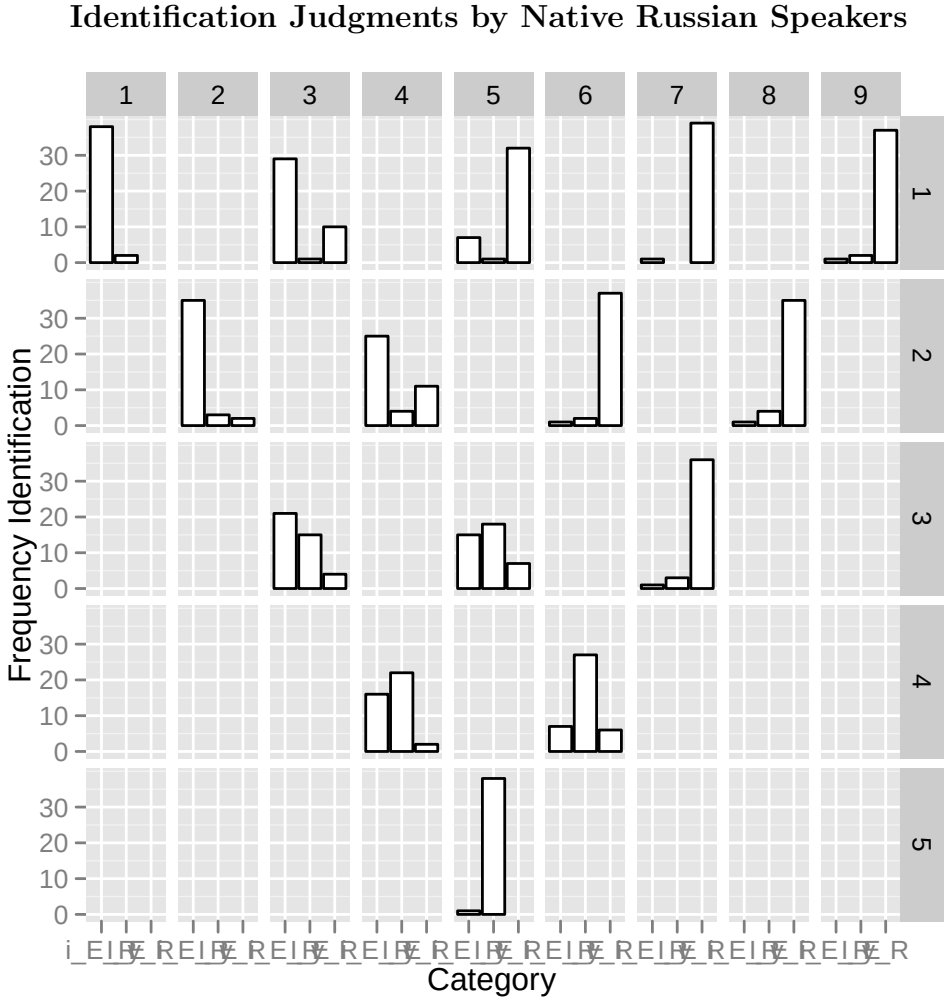


Figure G.10: Distribution of identification judgments by native Russian speakers for all stimuli in the 2D continuum. Each histogram represents, from left to right, classification as /i/, /ɪ/, and /ɨ/.

Stimulus	Item 1	Item 2	Nat_Eng	Beg	Int	Adv	Nat_Rus
11S	11	11	0.984	0.954	1	0.99	0.87
13S	13	13	0.939	0.985	0.978	0.957	0.958
15S	15	15	0.877	0.938	0.979	0.926	0.792
17S	17	17	0.984	0.954	1	0.947	0.857
19S	19	19	0.954	0.923	0.957	0.979	0.913
22S	22	22	0.985	0.954	0.978	1	0.909
24S	24	24	0.938	0.923	0.979	0.969	0.826
26S	26	26	0.939	0.924	0.979	0.927	0.875
28S	28	28	0.985	0.939	0.935	0.947	0.826
33S	33	33	0.923	0.985	1	0.968	0.652
35S	35	35	0.954	0.939	0.957	0.958	0.875
37S	37	37	0.923	0.984	0.979	0.947	0.875
44S	44	44	0.955	1	1	0.979	0.909
46S	46	46	1	0.939	0.957	0.99	0.773
55S	55	55	0.984	0.955	1	0.979	0.87

Table G.1: This table contains accuracy scores for all SAME pairs in the discrimination experiment for speakers at all levels of Russian proficiency

Stimulus	Item 1	Item 2	Nat_Eng	Beg	Int	Adv	Nat_Rus
11RR	11	13	0.386	0.114	0.034	0.188	0.313
11DR	11	22	0.279	0.136	0.194	0.109	0.267
13RR	13	15	0.25	0.095	0.281	0.266	0.438
13DL	13	22	0.159	0.114	0.094	0.047	0.313
13DR	13	24	0.476	0.227	0.258	0.25	0.25
13DD	13	33	0.953	0.932	0.903	0.938	0.643
15RR	15	17	0.023	0.045	0.156	0.145	0.2
15DL	15	24	0.432	0.25	0.219	0.143	0.25
15DR	15	26	0.318	0.262	0.258	0.19	0.267
15DD	15	35	0.864	0.886	0.969	0.969	0.733
17RR	17	19	0.023	0.093	0.033	0.127	0.133
17DL	17	26	0.256	0.386	0.156	0.141	0.143
17DR	17	28	0.114	0.19	0.194	0.125	0.125
17DD	17	37	0.909	1	0.969	0.984	0.6
19DL	19	28	0.163	0.159	0.065	0.094	0
22RR	22	24	0.295	0.25	0.031	0.095	0.267
22DR	22	33	0.795	0.886	0.875	0.797	0.4
24RR	24	26	0.349	0.295	0.419	0.317	0.563
24DL	24	33	0.773	0.721	0.531	0.734	0.417
24DR	24	35	0.744	0.909	0.903	0.891	0.615
24DD	24	44	0.884	1	0.935	0.984	0.846
26RR	26	28	0.233	0.205	0.194	0.159	0.2
26DL	26	35	0.841	0.795	0.781	0.844	0.6
26DR	26	37	0.558	0.698	0.688	0.619	0.533
26DD	26	46	0.953	0.955	0.969	0.938	0.867
28DL	28	37	0.791	0.721	0.767	0.787	0.438
33RR	33	35	0.273	0.31	0.233	0.302	0.5
33DR	33	44	0.364	0.302	0.355	0.328	0.313
35RR	35	37	0.318	0.302	0.161	0.234	0.375
35DL	35	44	0.302	0.341	0.452	0.422	0.385
35DR	35	46	0.349	0.205	0.219	0.27	0.25
35DD	35	55	0.595	0.455	0.621	0.667	0.714
37DL	37	46	0.524	0.318	0.3	0.349	0.5
44RR	44	46	0.186	0.093	0.156	0.078	0.188
44DR	44	55	0.295	0.136	0.188	0.177	0.5
46DL	46	55	0.273	0.186	0.188	0.226	0.6

Table G.2: This table contains accuracy scores for all pairs that are 1 step apart in the discrimination experiment for speakers at all levels of Russian proficiency

Stimulus	Item 1	Item 2	Nat_Eng	Beg	Int	Adv	Nat_Rus
11RRRR	11	15	0.682	0.568	0.531	0.635	0.933
11DRRR	11	24	0.841	0.5	0.742	0.5	0.643
11DDRR	11	33	0.977	1	0.969	0.969	0.917
13RRRR	13	17	0.614	0.455	0.452	0.625	0.714
13DRRR	13	26	0.727	0.659	0.767	0.73	0.688
13DDRR	13	35	0.955	0.977	1	0.968	0.733
15RRRR	15	19	0.31	0.364	0.4	0.422	0.563
15DLLL	15	22	0.628	0.488	0.484	0.444	0.733
15DRRR	15	28	0.318	0.419	0.594	0.444	0.333
15DDLL	15	33	0.977	0.977	0.903	0.935	0.867
15DDRR	15	37	0.837	0.977	0.938	0.984	0.813
17DLLL	17	24	0.705	0.477	0.563	0.563	0.667
17DDLL	17	35	0.909	0.977	1	0.984	0.867
19DLLL	19	26	0.386	0.476	0.5	0.25	0.188
19DDLL	19	37	0.909	1	1	0.952	0.813
22RRRR	22	26	0.744	0.837	0.563	0.774	0.923
22DRRR	22	35	0.909	0.955	0.966	0.984	0.813
22DDRR	22	44	0.953	1	0.969	0.969	0.714
24RRRR	24	28	0.767	0.524	0.75	0.656	0.75
24DRRR	24	37	0.886	0.93	1	0.938	0.867
24DDRR	24	46	0.864	0.907	0.969	1	0.813
26DLLL	26	33	0.909	0.744	0.938	0.875	0.938
26DDLL	26	44	0.952	0.977	0.968	0.984	0.938
28DLLL	28	35	0.955	0.886	0.938	0.921	0.938
28DDLL	28	46	0.932	1	1	0.969	0.933
33RRRR	33	37	0.705	0.636	0.625	0.703	0.933
33DRRR	33	46	0.705	0.744	0.645	0.641	0.8
33DDRR	33	55	0.721	0.674	0.656	0.766	0.643
37DLLL	37	44	0.814	0.619	0.719	0.766	0.875
37DDLL	37	55	0.841	0.795	0.742	0.734	0.857

Table G.3: This table contains accuracy scores for all pairs that are 2 steps apart in the discrimination experiment for speakers at all levels of Russian proficiency

References

Aaltonen, O., Eerola, O., Hellström, A., Uusipaikka, E., & Lang, A. H. (1997). Perceptual magnet effect in the light of behavioral and psychophysiological data. *Journal of the Acoustical Society of America*, *101*(2), 1090-1105.

Abramson, A. S. (1961). Identification and discrimination of phonemic tones. *Journal of the Acoustical Society of America*, *33*, 842.

Allen, J. S., Miller, J. L., & DeSteno, D. (2003). Individual talker differences in voice-onset-time. *Journal of the Acoustical Society of America*, *113*, 544-552.

Anderson, J. A., Silverstein, J. W., Ritz, S. A., & Jones, R. S. (1977). Distinctive features, categorical perception, and probability learning: Some applications of a neural model. *Psychological Review*, *84*(5), 413-451.

Andruski, J. E., Blumstein, S. E., & Burton, M. (1994). The effect of subphonemic differences on lexical access. *Cognition*, *52*, 163-187.

Angeli, A., Davidoff, J., & Valentine, T. (2008). Face familiarity, distinctiveness, and categorical perception. *The Quarterly Journal of Experimental Psychology*, *61*(5), 690-707.

Barrios, S. (2013). *Similarity in L2 phonology* (Unpublished doctoral dissertation). University of Maryland.

Bastian, J., & Abramson, A. S. (1962). Identification and discrimination of phonemic vowel duration. *Journal of the Acoustical Society of America*, *34*, 743.

Bastian, J., Delattre, P. C., & Liberman, A. M. (1959). Silent interval as a cue for the distinction between stops and semivowels in medial position. *Journal of the Acoustical Society of America*, *31*, 1568.

Bastian, J., Eimas, P. D., & Liberman, A. M. (1961). Identification and discrimination of a phonemic contrast induced by silent interval. *Journal of the Acoustical Society of America*, *33*, 842.

Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society*, *53*, 370-418.

Beale, J. M., & Keil, F. C. (1995). Categorical effects in the perception of faces. *Cognition*, *57*, 217-239.

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), 1551-1561.

Best, C. T. (1994). Emergence of native-language influences. In J. C. Goodman & H. C. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (p. 167-224). Cambridge, MA: MIT Press.

Best, C. T. (1995). A direct realist view of cross-language speech perception. *Speech perception and linguistic experience: Issues in cross-language research*.

Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustical Society of America*, *109*(2), 775-794.

Best, C. T., McRoberts, G. W., & Sithole, N. M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(3), 345-360.

Best, C. T., & Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics*, *20*, 305-330.

Blumstein, S. E., Myers, E. B., & Rissman, J. (2005). The perception of voice onset time: An fMRI investigation of phonetic category structure. *Journal of Cognitive Neuroscience*, *17*(9), 1353-1366.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International*, *5*(9/10), 341-345.

Boersma, P., & Weenink, D. (2005). *Praat: Doing phonetics by computer*. (Version 4.3.01) [Computer program]. Retrieved from http://www.praat.org/.

Byrd, D. (1992). Preliminary results on speaker-dependent variation in the TIMIT database. *Journal of the Acoustical Society of America*, *92*, 593-596.

Calder, A. J., Young, A. W., Perrett, D. I., Etcoff, N. L., & Rowland, D. (1996). Categorical perception of morphed facial expressions. *Visual Cognition*, *3*(2), 81-117.

Carney, A. E., Widin, G. P., & Viemeister, N. F. (1977). Noncategorical perception of stop consonants differing in VOT. *Journal of the Acoustical Society of America*, *62*(4), 961-970.

Chang, C. B. (2011). Systemic drift of l1 vowels in novice l2 learners. In *International congress of phonetic sciences* (Vol. XVII, p. 428-431).

Chew, P. A. (2003). *A computational phonology of russian* (Unpublished doctoral dissertation). Jesus College, University of Oxford, Parkland, FL, USA.

Chistovich, L. A. (1960). Classification of rapidly repeated speech sounds. *Akus-*

ticheskij Zhurnal, *6*, 392-398. (Translated in Soviet Physics-Acoustics, New York, 1961, 6, 393-398)

Crawford, J. C., & Wang, W. S.-Y. (1960). Frequency studies of English consonants. *Language & Speech*, *3*, 131-139.

Damper, R. I., & Harnad, S. R. (2000). Neural network models of categorical perception. *Perception and Psychophysics*, *62*(4), 843-867.

Davidoff, J., Davies, I., & Roberson, D. (1999). Colour categories in a stone-age tribe. *Nature*, *398*, 203-204.

Diesch, E., Iverson, P., Kettermann, A., & Siebert, C. (1999). Measuring the perceptual magnet effect in the perception of /i/ by German listeners. *Psychological Research*, *62*, 1-19.

Eimas, P. D. (1963). The relation between identification and discrimination along speech and nonspeech continua. *Language & Speech*, *6*, 206-217.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, *171*(3968), 303-306.

Elman, J. L. (1979). Perceptual origins of the phoneme boundary effect and selective adaptation to speech: A signal detection theory analysis. *Journal of the Acoustical Society of America*, *65*, 190-207.

Espy-Wilson, C. Y., Boyce, S. E., Jackson, M., Narayanan, S., & Alwan, A. (2000). Acoustic modeling of american english /r/. *Journal of the Acoustical Society of America*, *108*, 343-356.

Fant, G. (1960). *Acoustic theory of speech production*. The Hague, The Netherlands: Mouton.

Feldman, N., Richter, C., Falk, J., & Jansen, A. (2013, October). Predicting listeners' perceptual biases using low-level speech features. In *Paper presented at the Northeast Computational Phonology Workshop (NECPhon 7)*. Boston, MA.

Feldman, N. H., Griffiths, T. L., & Morgan, J. L. (2009). The influence of categories on perception: Explaining the perceptual magnet effect as optimal statistical inference. *Psychological Review*, *116*(4), 752-782.

Flanagan, J. (1972). *Speech analysis, synthesis and perception*. Berlin: Springer-Verlab.

Flege, F. E., Bohn, O. S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, *25*, 437-470.

Flege, J. E. (1987). The production of 'new' and 'similar' phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics*, *15*, 47-65.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (p. 233-276). York Press.

Fok, A. (1979). The frequency of occurrence of speech sounds and tones in Cantonese. In R. Lord (Ed.), *Hong Kong language papers.* Hong Kong: Hong Kong University Press.

Forrest, K., Weismer, G., Milenkovic, P., & Dougall, R. N. (1988). Statistical analysis of word-initial voiceless obstruents: preliminary data. *Journal of the Acoustic Society of America*, *84*, 115-123.

Fox, R. A. (1989). Dynamic information in the identification and discrimination of vowels. *Phonetica*, *46*, 97-116.

Fry, B. (1947). The frequency of occurrence of speech sounds in Southern English archives. *Néerlandaises de Phonétique Expérimentale.*, *20*, 103-106.

Fry, D. B., Abramson, A. S., Eimas, P. D., & Liberman, A. M. (1962). The identification and discrimination of synthetic vowels. *Language and Speech*, *5*, 171-189.

Fujimura, O. (1962). Analysis of nasal consonants. *Journal of the Acoustical Society of America*, *34*, 1865-1875.

Garner, W. R. (1974). *The processing of information adn structure.* New York: Wiley.

Gimson, A. C. (1980). *An introduction to the pronunciation of English* (3rd ed.). London: Edward Arnold.

Goldstone, R. L., Lippa, Y., & Shiffrin, R. M. (2001). Altering object representations through category learning. *Cognition*, *78*, 27-43.

Gow, D. W. (2001). Assimilation and anticipation in continuous spoken word recognition. *Journal of Memory and Language*, *45*, 133-159.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics.* New York: Wiley.

Grieser, D., & Kuhl, P. K. (1989). Categorization of speech by infants: Support for speech-sound prototypes. *Developmental Psychology*, *25*(4), 577-588.

Griffith, B. C. (1958). *A study of the relation between phoneme labeling and discriminability in the perception of synthetic stop consonants.* (Unpublished doctoral dissertation). University of Connecticut.

Guenther, F. H., & Gjaja, M. N. (1996). The perceptual magnet effect as an emergent property of neural map formation. *Journal of the Acoustical Society of America*, *100*(2), 1111-1121.

Halle, M., & Jones, L. G. (1959). *The sound pattern of russian. A linguistic and*

acoustical investigation. With an excursus on the contextual variants of the Russian vowels by Lawrence G. Jones. Mouton.

Halle, P. A., Best, C. T., & Levitt, A. (1999). Phonetic vs. phonological influences on french listeners' perception of american english approximants. *Journal of Phonetics*, *27*, 281-306.

Harnad, S. (Ed.). (1987). *Categorical perception: The groundwork of cognition*. New York: Cambridge University Press.

Harris, K. S. (1958). Cues for the discrimination of American English fricatives in spoken syllables. *Language and Speech*, *1*, 1-7.

Harris, K. S., Bastian, J., & Liberman, A. M. (1961). Mimicry and the perception of a phonemic contrast induced by silent interval: Electromyographic an acoustic measures. *Journal of the Acoustical Society of America*, *33*, 842.

Hashi, M., Honda, K., & Westbury, J. R. (2003). Time-varying acoustic and articulatory characteristics of American English [[turned r]]: A cross-speaker study. *Journal of Phonetics*, *31*, 3-22.

Hautus, M. J. (1995). Corrections for extreme proportions and their biasing effects on estimated values of d. *Behavior Research Methods*, *27*(1), 46-51.

Hayes, B. (2009). *Introductory phonology*. Malden, MA ; Oxford: Wiley/Blackwell.

Hess, U., Adams, R., & Kleck, R. (2009). The categorical perception of emotions and traits. *Social Cognition*, *27*(2), 320-326.

Hickok, G., & Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Sciences*, *4*(131-138).

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

Hillenbrand, J. M. (2013). Static and dynamic approaches to vowel perception. In G. S. Morrison & P. F. Assmann (Eds.), *Vowel inherent spectral change* (Vol. 9, p. 9-3-). Berlin Heidelberg: Springer-Verlab.

Hintzman, D. (1990). Human learning and memory: Connections and dissociations. In *Annual review of psychology* (p. 109-139). Palo Alto, CA: Annual Reviews Inc.

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

Hughes, G. W., & Halle, M. (1956). Spectral properties of fricative consonants. *Journal of the Acoustical Society of America*, *28*, 303-310.

Huttenlocher, J., Hedges, L. V., & Vevea, J. L. (2000). Why do categories affect stimulus judgment? *Journal of Experimental Psychology: General*, *129*(2), 220-

241.

Iverson, P., Hazan, V., & Bannister, K. (2005, November). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching english /r/-/l/ to japanese adults. *Journal of the Acoustical Society of America*, *118*(5), 3267-3278.

Iverson, P., & Kuhl, P. K. (1995). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *Journal of the Acoustical Society of America*, *97*(1), 553-562.

Iverson, P., & Kuhl, P. K. (1996). Influences of phonetic identification and category goodness on American listeners' perception of /r/ and /l/. *Journal of the Acoustical Society of America*, *99*(2), 1130-1140.

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

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, *87*, B47-B57.

Jacewicz, E., & Fox, R. A. (2013). Cross-dialectal differences in dynamic formant patterns in american english vowels. In G. S. Morrison & P. F. Assmann (Eds.), *Vowel inherent spectral change, modern acoustics and signal processing* (p. 177-198). Berlin Heidelberg: Springer-Verlab.

Jassem, W. (1962). Noise spectra of swedish, english, and polish fricatives. In *Proceedings of the sepech communication seminar* (p. 1-4).

Jeffress, L. A. (1948). A place theory of sound localization. *Journal of Computational Physiological Psychology*, *41*, 35-39.

Jenkins, J. J., Strange, W., & Trent, S. A. (1999). Context-independent dynamic information for the perception of coarticulated vowels. *Journal of the Acoustical Society of America*, *106*, 438-448.

Johannesson, M., & Lundagard, K. (2001). *The problem of combining integral and separable dimensions* (Tech. Rep.). Lund University.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of english as a second language. *Cognitive Psychology*, *21*, 60-99.

Jongman, A., Wayland, R., & Wong, S. (2000). Acoustic characteristics of English fricatives. *Journal of the Acoustical Society of America*, *108(3)*, 1252-1263.

Joos, M. (1948). *Acoustic phonetics* (Vol. 23). Baltimore: Linguistics Society of America.

Kewley-Port, D. (1983). Time-varying features as correlates of place of articulation in stop consonants. *Journal of the Acoustical Society of America*, *73*, 322-335.

Kleinschmidt, D. F., & Jaeger, T. F. (2011). A Bayesian belief updating model of phonetic recalibration and selective adaptation. In *2nd acl workshop on cognitive modeling and computational linguistics.*

Kleinschmidt, D. F., & Jaeger, T. F. (2012). A continuum of phonetic adaptation: Evaluating an incremental belief-updating model of recalibration and selective adaptation. In *Proceedings of the 34th annual conference of the cognitive science society.* Sapporo, Japan.

Kozhevnikov, V. A., & Chistovich, L. A. (1965). *Rech' artikuljatsia i vosprijatije.* Moscow-Leningrad. (Translated in Speech: Articulation and perception, Washington: Joint Publications Research Service, 1966, 30, 543)

Kröger, B. J., Birkholz, P., Kannampuzha, J., & Neuschaefer-Rube, C. (2007). Modeling the perceptual magnet effect and categorical perception using self-organizing neural networks. *International Congress of Phonetic Sciences*, 789-792.

Kronrod, Y., Barrios, S., Winn, M., Idsardi, W., & Feldman, N. (2014). Relationships between phoneme identification and discrimination change with varying L2 proficiency. In *Selected proceedings of the second language research forum.* Somerville, MA: Cascadilla Press.

Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception and Psychophysics*, *50*(2), 93-107.

Kuhl, P. K. (1993a). Early linguistic experience and phonetic perception: Implications for theories of developmental speech perception. *Journal of Phonetics*, *21*, 125-139.

Kuhl, P. K. (1993b). Innate predispositions and the effects of experience in speech perception: The native language magnet theory. In B. de Boysson-Bardies (Ed.), *Developmental neurocognition: Speech and face processing in the first year of life* (p. 259-274). The Hague: Kluwer Academic Publishers.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, *255*(5044), 606-608.

Lacerda, F. (1995). The perceptual-magnet effect: An emergent consequence of exemplar-based phonetic memory. In K. Elenius & P. Branderud (Eds.), *Proceedings of the XIIIth International Congress of Phonetic Sciences* (Vol. 2, p. 140-147). Stockholm: KTH and Stockholm University.

Lado, R. (1957). *Linguistics across cultures: Applied linguistics for language teachers.* Ann Arbor, MI: University of Michigan Press.

Lago, S., Kronrod, Y., Scharinger, M., & Idsardi, W. (2010). Categorical perception of [s] and [sh]: An MMN study.. Poster presented at the Neurobiology of Language Conference. San Diego, CA.

Lago, S., Scharinger, M., Kronrod, Y., & Idsardi, W. J. (Submitted). Categorical effects in fricative perception are reflected in cortical source information. *Brain and Language*.

Larkey, L. S., Wald, J., & Strange, W. (1978). Perception of synthetic nasal consonants in initial and final syllable position. *Perception and Psychophysics*, *23*(4), 299-312.

Lenneberg, E. (1967). *Biological foundations of language*. New York: Wiley.

Leontev, A. A. (1961). I. A. Bouden de Kurtene i Peterburgskaia shkola Russkoi lingvistiki. *Voprosy iazykoznaniia*, *4*.

Liberman, A. M., Cooper, F., Shankweiler, D., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*, 431-461.

Liberman, A. M., Delattre, P. C., Cooper, F. S., & Gerstman, L. J. (1954). The role of consonant-vowel transitions in the perception of the stop and nasal consonants. *Psychological Monographs*, *68*(8 (379)).

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*(5), 358-368.

Liberman, A. M., Harris, K. S., Kinney, J. A., & Lane, H. (1961). The discrimination of relative onset-time of the components of certain speech and nonspeech patterns. *Journal of Experimental Psychology*, *61*(5), 379-388.

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

Lisker, L., & Abramson, A. S. (1964b, August). Stop categories and voice onset time. In *Proceedings of the fifth international congress of phonetic sciences.* Munster.

Lively, S. E., & Pisoni, D. B. (1997). On prototypes and phonetic categories: A critical assessment of the perceptual magnet effect in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, *23*(6), 1665-1679.

Lotto, A. J., & Holt, L. L. (2011). Psychology of auditory perception. *WIREs Cognitive Science*, *2*, 479-489.

Lotto, A. J., Kluender, K. R., & Holt, L. L. (1998). Depolarizing the perceptual magnet effect. *Journal of the Acoustical Society of America*, *103*(6), 3648-3655.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd edition ed.). Taylor and Francis.

Macmillan, N. A., Kaplan, H. L., & Creelman, C. D. (1977). The psychophysics of categorical perception. *Psychological Review*, *84*, 452-471.

Marr, D. (1982). *Vision: A computational investigation in the human representation of visual information*. San Francisco: Freeman & Co.

Massaro, D. W. (1987a). Categorical partition: A fuzzy logical model of categorical behavior. In S. Harnad (Ed.), *Categorical perception: the groundwork of cognition* (p. 254-283). Cambridge University Press.

Massaro, D. W. (Ed.). (1987b). *Speech perception by ear and eye.* London: Lawrence Erlbaum Associates.

Massaro, D. W., & Oden, G. C. (1980). Evaluation and integration of acoustic features in speech perception. *The Journal of the Acoustical Society of America*, *67*, 996-1013.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*, B101-B111.

Mayo, C., & Turk, A. (2005). The influence of spectral distinctiveness on acoustic ue weighing in children's and adults' speech perception. *Journal of the Acoustic Society of America*, *118*(3), 1730-1741.

McClelland, J. L. (2009). The place of modeling in cognitive science. *Topics in Cognitive Science*, *1*, 11-38.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*, 1-86.

McMurray, B. (2009). *Klattworks: A [somewhat] new systematic approach to formant-based speech synthesis for empirical research.* Manuscript in preparation.

McMurray, B., Aslin, R. N., & Toscano, J. C. (2009). Statistical learning of phonetic categories: insights from a computational approach. *Developmental Science*, *12*(3), 369-378.

McQueen, J. M., Jesse, A., & Norris, D. (2009). No lexical–prelexical feedback during speech perception or: Is it time to stop playing those Christmas tapes? *Journal of Memory and Language*, *61*, 1-18.

Miller, J. L. (1981). Effects of speaking rate on segmental distinctions. In P. D. Eimas & J. L. Miller (Eds.), *Perspectives on the study of speech* (p. 39-74). Hillsdale, NJ: Erlbaum.

Miller, J. L. (1994). On the internal structure of phonetic categories: A progress report. *Cognition*, *50*, 271-285.

Miller, J. L., & Eimas, P. D. (1977, March). Studies on the perception of place and manner of articulation: A comparison of the labial-alveolar and nasal-stop distinctions. *Journal of the Acoustical Society of America*, *61*(3), 835-845.

Miller, J. L., & Volaitis, L. E. (1989). Effect of speaking rate on the perceptual structure of a phonetic category. *Perception and Psychophysics*, *46*, 505-512.

Minagawa-Kawai, Y., Mori, K., & Sato, Y. (2005). Different brain strategies underlie the categorical perception of foreign and native phonemes. *Journal of Cognitive Neuroscience*, *17*(9), 1376-1385.

Mines, M. A., Hanson, B. F., & Shoup, J. E. (1978). Frequency of occurrence of phonemes in conversational English. *Language & Speech*, *21*(3), 221-241.

Nearey, T. M. (1997). Speech perception as pattern recognition. *Journal of the Acoustical Society of America*, *101*, 3241-3256.

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*, 1297-1308.

Nearey, T. M., & Hogan, J. T. (1986). Phonological contrast in experimental phonetics: Relating distributions of production data to perceptual categorization curves. In J. J. Ohala & J. J. Jaeger (Eds.), *Experimental phonology* (p. 141-162). Orlando, FL: Academic Press.

Newell, A., & Simon, H. (1976). Computer science as empirical inquiry: Symbols and search. *Communications of ACM*, *19*, 113-126.

Newman, R. S., Clouse, S. A., & Burnham, J. L. (2001). The perceptual consequences of within-talker variability in fricative production. *Journal of the Acoustical Society of America*, *109*, 1181-1196.

Obleser, J., Leaver, A., VanMeter, J., & Rauschecker, J. P. (2010). Segregation of vowels and consonants in human auditory cortex: Evidence for distributed hierarchical organization. *Frontiers in Psychology*, *1*(232).

Oden, G. C., & Massaro, D. W. (1978). Integration of feature information in speech perception. *Psychological Review*, *85*, 172-191.

Ohde, R. N., & German, S. R. (2011, September). Formant onsets and formant transitions as developmental cues to vowel perception. *Journal of the Acoustical Society of America*, *130*(3), 1628-1642.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The edinburgh inventory. *Neuropsychologia*, *9*, 97-113.

Penfield, W., & Roberts, L. (1959). *Speech and brain mechanisms*. Princeton, NJ: Princeton University Press.

Perez, C. A., Engineer, C. T., Jakkamsetti, V., Carraway, R. S., Perry, M. S., & Kilgard, M. P. (2013). Different timescales for the neural coding of consonant and vowel sounds. *Cerebral Cortex*, *23*(3), 670-683.

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the

vowels. *Journal of the Acoustical Society of America*, *24*(2), 175-184.

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

Pisoni, D. B. (1975). Auditory short-term memory and vowel perception. *Memory and Cognition*, *3*(1), 7-18.

Pisoni, D. B. (1977). Identification and discrimination of the relative onset time of two component tones: Implications for voicing perception. *Journal of the Acoustical Society of America*, *61*(5), 1352-1361.

Pisoni, D. B., & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception and Psychophysics*, *15*(2), 285-290.

Poeppel, D., Idsardi, W. J., & van Wassenhove, V. (2008). Speech perception at the interface of neurobiology and linguistics. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*, 1071-1086.

Polka, L., & Bohn, O.-S. (2011, October). Natural referent vowel (nrv) framework: An emerging view of early phonetic development. *Journal of Phonetics*, *39*(4), 467-478.

Polka, L., & Strange, W. (1985). Perceptual equivalence of acoustic cues that differentiate /r/ and /l/. *Journal of the Acoustical Society of America*, *78*(4), 1187-1197.

Rato, A. (2014). Effects of perceptual training on the identification of English vowels by native speakers of European Portuguese. In *Proceedings of the international symposium on the acquisition of second language speech* (Vol. 5, p. 529-546). Concordia Working Papers in Applied Linguistics.

Repp, B. H. (1981). Two strategies in fricative discrimination. *Perception and Psychophysics*, *30 (3)*, 217-227.

Repp, B. H., Healy, A. F., & Crowder, R. G. (1979). Categories and context in the perception of isolated steady-state vowels. *Journal of Experimental Psychology: Human Perception and Performance*, *5*(1), 129-145.

Roberson, D., Hanley, J. R., & Pak, H. (2009). Thresholds for color discrimination in english and korean speakers. *Cognition*, *112*, 482-487.

Salminen, N. H., Titinen, H., & May, P. J. C. (2009). Modeling the categorical perception of speech sounds: A step toward biological plausibility. *Cognitive, Affective, and Behavioral Neuroscience*, *9*(3), 304-313.

Sauter, D., LeGuen, O., & Huan, D. (2011). Categorical perception of emotional facial expressions does not require lexical categories. *Emotion*, *11*(6).

Schwarz, G. E. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*(2), 461-464.

Selkirk, E. O. (1984). On the major class features and syllable theory. In M. Aronoff & R. T. Oerhle (Eds.), *Language sound structure: Studies in phonology dedicated to Morris Halle by his teacher and students.* (p. 107-113). MIT Press.

Shi, L., Griffiths, T. L., Feldman, N. H., & Sanborn, A. N. (2010). Exemplar models as a mechanism for performing bayesian inference. *Psychonomic Bulletin & Review*, *14*(4), 443-464.

Silbert, N. H., Townsend, J. T., & Lentz, J. J. (2009). Independence and separability in the perception of complex nonspeech sounds. *Attention, Perception, & Psychophysics*, *71*(8), 1900-1915.

Sporer, S. L. (2001). Recognizing faces of other ethnic groups: An integration of theories. *Psychology, Public Policy, and Law*, *7*(1), 36-97.

Stevens, K. N. (1966, August). On the relations between speech movements and speech perception. In *Meeting of the XVIII international congress of psychology.* Moscow. (Zeitshrift fur Phonetik, Sprachwissensschaft und Kommunikationsforschung)

Stevens, K. N., & Blumstein, S. E. (1978). Invariant cues for place of articulation in stop consonants. *Journal of the Acoustical Society of America*, *64*, 1358-1368.

Stevens, K. N., Ohman, S. E. G., & Liberman, A. M. (1963). Identification and discrimination of rounded and unrounded vowels. *Journal of the Acoustical Society of America*, *35*, 1900.

Stevens, K. N., Ohman, S. E. G., Studdert-Kennedy, M., & Liberman, A. M. (1964). Cross-linguistic study of vowel discrimination. *Journal of the Acoustical Society of America*, *36*, 1989.

Stevens, S. S., Volkmann, J., & Newman, E. B. (1937). A scale for the measurement of the psychological magnitude pitch. *Journal of the Acoustical Society of America*, *8*, 185-190.

Strange, W. (1989, May). Evolving theories of vowel perception. *Journal of the Acoustical Society of America*, *85*(5), 2081-2087.

Strevens, P. (1960). Spectra of fricative noise in human speech. *Language and Speech*, *3*, 32-49.

Studdert-Kennedy, M., Liberman, A. M., & Stevens, K. N. (1963). Reaction time to synthetic stop consonants and vowels at phoneme centers and at phoneme boundaries. *Journal of the Acoustical Society of America*, *35*, 1900.

Studdert-Kennedy, M., Liberman, A. M., & Stevens, K. N. (1964). Reaction time during the discrimination of synthetic stop consonants. *Journal of the Acoustical Society of America*, *36*, 1989.

Sun, R. (2008). Introduction to computational cognitive modeling. In R. Sun (Ed.), *Cambridge handbook of computational psychology* (p. 3-19). New York: Cambridge

University Press.

Sun, R., Coward, A., & Zenzen, M. (2005). On levels of cognitive modeling. *Philosophical Psychology*, *18*(5), 613-637.

Sussman, H. A., McCaffrey, H. A., & Matthews, S. A. (1991). An investigation of locus equations as a source of relational invariance for stop place categorization. *Journal of the Acoustical Society of America*, *90*, 1309-1325.

Sussman, J. E., & Gekas, B. (1997). Phonetic category structure of [I]: Extent, best exemplars, and organization. *Journal of Speech, Language, and Hearing Research*, *40*, 1406-1424.

Theodore, R. M., Miller, J. L., & DeSteno, D. (2009, June). Individual talker differences in voice-onset-time: Contextual influences. *Journal of the Acoustical Society of America*, *125*(6), 3974-3982.

Theodore, R. M., Myers, E. B., & Lomibao, J. (2013). Listeners sensitivity to talker differences in voice-onset-time: Phonetic boundaries and internal category structure. In *Proceedings of meetings on acoustics (poma) 19*.

Thorburn, W. M. (1915). Occam's razor. *Mind*, *24*, 287-288.

Thorburn, W. M. (1918). The myth of occam's razor. *Mind*, *27*, 345-353.

Thyer, N., Hickson, L., & Dodd, B. (2000). The perceptual magnet effect in Australian English vowels. *Perception and Psychophysics*, *62*(1), 1-20.

Tomaschek, F., Truckenbrodt, H., & Hertrich, I. (2011). Processing german vowel quantity: Categorical perception or perceptual magnet effect? *Proceedings of the 17th International Conference of Phonetic Sciences*, 2002-2005.

Toscano, J. C., & McMurray, B. (2010). Cue integration with categories: Weighting acoustic cues in speech using unsupervised learning and distributional statistics. *Cognitive Science*, *34*, 434-464.

Treisman, M., Faulkner, A., Naish, P. L. N., & Rosner, B. S. (1995). Voice-onset time and tone-onset time: The role of criterion-setting mechanisms in categorical perception. *Quarterly Journal of Experimental Psychology*, *48A*(334-366).

Treisman, M., & Williams, T. C. (1984). A theory of criterion setting with an application to sequential dependencies. *Psychological Review*, *91*, 68-111.

Valentine, T., & Endo, M. (1992). Towards an exemplar model of face processing: The effects of race and distinctiveness. *Quarterly Journal of Experimental Psychology*, *44A*, 671-703.

Vallabha, G. K., & McClelland, J. L. (2007). Success and failure of new speech category learning in adulthood: Consequences of learned Hebbian attractors in topographic maps. *Cognitive, Affective, and Behavioral Neuroscience*, *7*, 53-73.

Vallabha, G. K., McClelland, J. L., Pons, F., Werker, J. F., & Amano, S. (2007).

Unsupervised learning of vowel categories from infant-directed speech. *Proceedings of the National Academy of Sciences*, *104*, 13273-13278.

Werker, J. F., Gilbert, J. H. V., Humphrey, K., & Tees, R. C. (1981). Developmental aspects of crosslanguage speech perception. *Child Development*, *52*, 349-355.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*, 49-63.

Winn, M., & Litovsky, R. (Submitted). Testing functional spectral resolution for speech sounds in listeners with cochlear implants. *Journal of the Acoustical Society of America*.

Wioland, F. (1972). Estimation de la frequence des phonemes en francais parle. *Travaux de l'Institut de Phonetique*, *4*, 177-204.

Wood, C. C. (1976). Discriminability, response bias, and phoneme categories in discrimination of voice onset time. *Journal of the Acoustic Society of America*, *60 (6)*, 1381-1389.

Yingyong, Q., & Fox, R. A. (1992, March). Analysis of nasal consonants using perceptual linear prediction. *Journal of the Acoustical Society of America*, *91*(3).

Zue, V. W., & Laferriere, M. (1979). Acoustic study of medial /t,d/ in American English. *Journal of the Acoustical Society of America*, *66*, 1039-1050.

Zwicker, E. (1961, February). Subdivision of the audible frequency range into critical bands. *The Journal of the Acoustical Society of America*, *33*, 248.