ABSTRACT

Title of Dissertation: LANGUAGE SCIENCE MEETS COGNITIVE SCIENCE: CATEGORIZATION AND ADAPTATION

Christopher Cullen Heffner, Doctor of Philosophy, 2017

Dissertation directed by: Professor Rochelle S. Newman
Department of Hearing and Speech Sciences

Professor William J. Idsardi
Department of Linguistics

Questions of domain-generality—the extent to which multiple cognitive functions are represented and processed in the same manner—are common topics of discussion in cognitive science, particularly within the realm of language. In the present dissertation, I examine the domain-specificity of two processes in speech perception: category learning and rate adaptation. With regard to category learning, I probed the acquisition of categories of German fricatives by English and German native speakers, finding a bias in both groups towards quicker acquisition of non-disjunctive categories than their disjunctive counterparts. However, a study using an analogous continuum of non-speech sounds, in this case spectrally-rotated musical instrument sounds, did not show such a bias, suggesting that at least some attributes of the phonetic category learning process are unique to speech. For rate adaptation, meanwhile, I first report a study examining rate adaptation in Modern Standard Arabic (MSA), where consonant length is a contrastive part of the phonology; that is, where words can be distinguished from one another by the length of the consonants that make them up. I found that changing the rate of the beginning of a sentence can lead a consonant towards the end of the sentence to change in its perceived duration; a short consonant can sound like a long one, and a long consonant can sound like a short one. An analogous experiment examined rate adaptation in event segmentation, where adaptation-like effects had not previously been explored, using recordings of an actor interacting with a touchscreen. I found that the perception of actions can also be affected by the rate of previously-occurring actions. Listeners adapt to the rate at the beginning of a series of actions when deciding what they saw last in that series of actions. This suggests that rate adaptation follows similar lines across both domains. All told, this dissertation leads to a picture of domain-specificity in which both domain-general and domain-specific processes can operate, with domain-specific processes can help scaffold the use of domain-general processing.

LANGUAGE SCIENCE MEETS COGNITIVE SCIENCE: CATEGORIZATION AND ADAPTATION

by

Christopher Cullen Heffner

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
2017

Advisory Committee:
Professor Rochelle S. Newman, Co-Chair
Professor William J. Idsardi, Co-Chair
Professor Catherine E. Carr, Graduate Dean’s Representative
Professor Ellen F. Lau
Professor L. Robert Slevc

Dedication

For my parents

Acknowledgements

Getting a doctorate requires a lot of support. I'm so lucky to have been able to experience the wonderful community of the University of Maryland. In fact, thanks to being split across two departments, a program, and a center, I've gotten to enjoy it many times over! My list of acknowledgements is correspondingly long, and I apologize in advance for any inadvertent oversights.

Thanks go first of all to my advisors, Rochelle Newman and Bill Idsardi. Forms often assume that PhD students have one advisor, or, at the very least, that they have one they can nominate as their "main" advisor. For me, that has never been the case. Bill and Rochelle have been an amazing pair, working together seamlessly to make sure I can develop into the best scientist that I can be. Whether it has been Rochelle sending me meticulously-edited manuscripts at all times of the day, or Bill making sure again and again that I'm attending to the big-picture meaning of my experiments, they have done nothing but encourage and inspire me over the course of my five years here, and done so as a team. How they managed to do that while also chairing their respective departments is beyond me, but that is a testament to their amazing dedication to making their communities a better place. I feel like I'll have a lot to live up to when being a mentor in the future; but I'll also have two excellent examples of what mentorship should be.

Having an office next to Ellen Lau has been one of the most pleasant experiences of my life. If I'm having a bad day, all I need to do is wait for her to have a meeting with her office door open, and her infectious laughter will brighten my day. Her sunny personality is matched by an amazing depth of knowledge on anything cognitive neuroscience. Although sounds aren't her usual wheelhouse, I've been amazed by the insights she's been able to bring to the projects that have interested me. And, of course, it's always been pleasant to have a Spartan around at all times. Go Green!

When I first sent Catherine Carr an email just to broach the idea of talking about having her join my committee, I was expecting a measured response; after all, language science isn't always closely connected to neuroethology. I cannot say I was expecting her to immediately and enthusiastically say "yes!" Her curiosity knows no bounds; every day of her neurobiology of hearing class certainly opened *my* eyes to new insights, and the pleasure I got from it was further enhanced by seeing how much delight *she* took from learning, too. The NACS program is lucky to have her, and I've been fortunate to have her contribute her insights on my committee.

Is there any topic vaguely related to music cognition that Bob Slevc is not interested in? I don't think anyone has found the outer limit yet. It was always a plan of mine to find *some* way to collaborate with Bob on a project, so it's been fantastic to have those plans realized in spades. Teaching musicians about German fricatives? Sure! Taking a term paper and turning it into some published speculation about prosody and music? Why not! Working with Bob, and having him on my committee, have been nothing but pleasant. We still need to grab a drink at Town Hall sometime.

I've also had other collaborators at UMD. At my first Language Science Day, Rochelle told me to go talk to "the guy in the pink shirt" because she thought I might want to collaborate with him; it turned out Anton Rytting was doing research with Arabic and might be interested in some of my rate adaptation experiments. I am so grateful to Anton and Buthainah Al-Thowaini for their help with the Arabic project presented in Chapter 3.2. Buthianah in particular did the yeoman's work of running all of the Arabic-speaking participants, which is incredible, and had some very helpful edits in Chapter 3.2. I owe Matt Goupell and Brittany Jaekel a debt of gratitude for indulging my interest in word segmentation in an interesting population, and putting up with my workload over the past year. Jared Novick is a pleasure to collaborate with, even when every experiment you try with him happens to fail miserably. Stupid psychic participants.

Through my advisors, I have had an amazing set of academic siblings, both through Rochelle and through Bill. On the Rochelle side of my family tree, Giovanna Morini is one of the kindest people I know; she had to be to put up with my Spanglish (siempre con acentos agudos adecuados). I can't wait to see what amazing things she'll do with her own lab! Amritha Mallikarjun is a wonderful little sister. I'm glad she joins with me to oppose any attempt to turn Tri-Puppies into a hurting fraternity. I've always been so impressed with Melissa Stockbridge for doing the joint MA/PhD route. We've also had excellent lab managers. Quite frankly, I'm surprised but glad Chantal Hoff did not immediately quit her job after Nuns on the Run. Emily Shroads has helped make the last years of my PhD smooth sailing.

Bill's lab is a bit more spread out, but is not less amazing for being what it is. Ewan Dunbar and Shannon Barrios provided support and encouragement at the beginning of my PhD career. Shannon in particular must be commended for linking me up with the apartment I've lived in for the past four years. I'm proud to be Bill's fourth consecutive student in this place! Rachael Richardson knows the importance of rate adaptation. Chris Neufeld has been an excellent officemate; I'm so glad to always have someone to talk about sounds, squiggles, and Canadian politics with. I wish Max Papillon could also be there to have a three-way conversation. Mike Key served as a great mentor when I was TAing, and hooked me up with his German fricatives to boot.

I have been able to spend time with some amazing people as a part of my program. I remember the first time I met Molly Hyer, before my first semester at UMD, I thought to myself, "this woman will be the queen of NACS". I was… entirely right. Thanks for making so many memories (and sometimes losing them) over the years. I was also entirely unsurprised that Alix Kowalski and I knew someone in common (shout out to Andy Webb), given that she was from Royal Oak. It's been so wonderful to have a Michigander to talk to. (Pro tip: have Alix introduce you to her dad. He's amazing.) Zoe Schlueter, Eric Pelzl, Alia Lancaster, Rachel Adler, and the other members of our board game support group as well as my NACS cohort (besides Molly, Sarah Blankenship, Amanda Burton, Farzad Ehtemam, Ronny Gentry, Greg Perrin, Clare Sengupta, and Jared Shamwell) have helped keep me as sane as I'll ever be.

Besides the people mentioned above, there have been many other students who have made my life better. From NACS, I'd like to thank Erika Hussey, Alice Jackson, Nuria AbdulSaber, Katie Willis, Susan Tuebner-Rhodes, Dan Bryden, Adam Jones, Amanda Chicoli, Jeff Chrabaszcz, Andrew Venezia, Alex Presacco, Matt Swierzbinski, Kathryn Yoo, Krystyna Solarana, Aminah Sheikh, Felix Bartsch, Andrew Borrell, Francisco Cervantes, Matt Coon, Jess Ellis, Adam Fishbein, Rui Hu, Graham Marquart, Dustin Moraczewski, Mattson Ogg, Zoe Ovans, Kevin Schneider, and Hetal Shah. From LING, thanks go to Shevaun Lewis, Brooke Larson, Dave Kush, Michael Gagnon, Wing Yee Chow, Alexis Wellwood, Megan Sutton, Dan Parker, Sol Lago, Aaron Steven White, Naho Orita, Angela He, Kaitlyn Harrigan, Dustin Chacón, Carolina Peterson, Shota

Momma, Julie Gerard, Rachel Dudley, Lara Ehrenhofer, Allyson Ettinger, Mike Fetters, Phoebe Gaston, Jeff Green, Mina Hirzel, Kasia Hitczenko, Nick Huang, Paulina Lyskawa, Anton Malko, Gesoel Mendes, Dongwoo Park, Laurel Perkins, Rodrigo Ranero, and Zack Stone. A special thank you goes to Tara Mease for being an excellent hand model and to Joseph Pattisall for shooting the videos used in the event segmentation. Thanks also to Anna Namyst, Elizabeth Nguyen, and Natalia Lapinskaya for allowing me to make horrible mistakes with fancy gadgets, and to the many Baggett fellows throughout the years for their support for LING in general and my research in particular. And, within HESP, thanks to Julie Cohen, Michelle Erskine, Julianne Gabarino, Allison Johnson, Maureen Shader, Arifi Waked, Jess Wess, and Nina Hsu.

I have also benefited from excellent undergraduates and students to help me conduct my research. The Language Development Lab has probably had too many alumni during the five years I've been here to enumerate, but they have all been excellent! I have also directly overseen researchers who have supported either the research described in this dissertation or a closely allied study. Stephen DeVilbiss, James Gray, Joshua Weaver, Daniela Vazquez, Sanna Darwish, Scott Kaplowitz, Peter Deaville, Garrett Yocklin, Nichole Mgboji, and Michelle Mei have all been extremely helpful in making these projects a reality.

Plenty of faculty and staff have also supported my development. We all owe Colin Phillips a debt of thanks for shepherding the Language Science Center. The insane commitment he has to the cause is incredible to witness. I doubt I will ever encounter anyone else in my entire life who can explain computational theories of language as comprehensibly as Naomi Feldman. Norbert Hornstein and Howard Lasnik have immense stores of wisdom that I've been grateful to tap into. Juan Uriagereka has been a great partner in crime for our biolinguistics class, with ample support from Peggy Antonisse and Tonia Bleam. Nan Bernstein Ratner made the most of a one-minute question period to inspire a new line of research. Jan Edwards always has something interesting to say; I wish we could have overlapped at UMD for longer. I've also benefitted from an amazing group of support staff. Pam Komarek is a personal hero, and her and the other NACS staff (Emerald Brooks, Gaëlle Kolb, and Blaze Buck) are essential for keeping the program ticking. Kim Kwok, Kathy Faulkingham, and Denise Best as well as Valerie Brownlee and Rachelle Beasley do the same for LING and HESP, respectively.

I have been supported by a variety of funding sources. The research directly reported here was funded in part from a Dean's Research Initiative grant from the College of Behavioral and Social Sciences at UMD and in part from a Doctoral Dissertation Research Improvement (DDRI) grant from National Science Foundation (Grant BCS1650791). I as a researcher have been funded by a National Science Foundation Graduate Research Fellowship, a Flagship Fellowship from the Graduate School of the University of Maryland, and a William Hodos Dissertation Assistantship from NACS. Although I was not directly funded off either grant, I have also benefited immensely from the community (and some of the support mechanisms) engendered by the IGERT (0801465) and NRT (1449815) grants to UMD from the NSF.

My advisors have also been kind enough to allow me to gallivant across states and even oceans to get research done. I spent two weeks in Tübingen, Germany at the University of Tübingen to collect data for the experiment described in Chapter 2.2, hosted in the lab of Andrea Weber. Andrea, Ann-Kathrin Grohe, and Sara Beck were generous in both letting me use their lab facilities and creating a welcoming and supportive environment while I was there. Sarah Schwarz and Lisa Kienzle were extremely responsible research assistants. I was also fortunate enough to overlap with Holger Mitterer and Eva Reinisch while I was there, which was a pleasure. I also spent two months in somewhat-less-exotic Columbus, Ohio gathering data on another category learning project. Thanks there goes to Laura Wagner for allowing me to be a part of the amazing language lab at the Center Of Science and Industry, and to Rachel Fields, who, despite having to come late, more than made up for it in her work ethic. My Oddfellows trivia team deserves a shout-out for making my stay fun. I've also been extremely fortunate to have something to look forward to while I've been writing this thing. To the folks at UConn, especially Emily Myers, I can't wait to start! A small portion of Chapter 4 represents thinking that went into a grant proposal of ours.

I have often wondered why people writing dissertations often neglect to include people in their lives *before* their PhD program. I wouldn't be where I am without the educators who inspired me before I ever set foot in College Park. From Michigan State, I will always be grateful to Laura Dilley, Karthik Durvasula, and Cristina Schmitt for teaching me about the joys of scientific research. Bess German is the one who got me to MSU in the first place; she saw what a good fit the university and I would be for each other, and made sure I got there. I'm still fortunate enough to be collaborating with Liz Wieland and Soo-Eun Chang, even though the latter moved to a certain *other* university in Michigan. I also had great group of teachers during my primary and secondary education. I likely wouldn't be writing this dissertation if Kevin Johnson hadn't told me to read Steven Pinker's books, transforming my hobby of linguistics into a lifelong passion. Dave Motes taught me that writing can be both terrifying and exhilarating. Clarice Hagen helped me channel my nerdiness in the pursuit of trivia glory. When my life took a suddenly, sharp turn for the worse in fourth grade, Martha Perbix helped turn it back around. And of course, to all my Wayzata and Michigan State friends, thank you.

I've been able to have a few outlets outside of academia to help keep myself grounded. My church family at the Episcopal Church of the Ascension in Silver Spring has been marvelous. Many thanks to the rector, Joan, and deacon, Terri, for their leadership. Thanks also to the members of St. Ælred's Young Adults group, especially Emily Clough, Darah Wright, and Arabella Littlepage for their time as co-leaders of the group with me. I'm looking forward to another bad movie night! If you had told me in 2012 that one of my closest groups of friends during my graduate career would be a group of Redditors, I'd be shocked. My trivia team, the Masterpieces of Manflesh, has been jam-packed with amazing people. In a city that was a culture shock to my Midwestern roots, I can't believe such an amiable group could exist; although I can only highlight a few, I'm grateful to all of them. Alvin Chen was the very first person to show up on that very first Tuesday, and has been there ever since. He's the heart of the group, and one of the most generous people I know. Rob Blackwell has always been up for Michigan bonding time at Ivy & Coney with me and Alix. Jeff Naber does the best Minnesotan impression of any non-Minnesotan I've known (tied with Karthik). The esteemed Calvert Hills Board Game Association has also provided me with meaningful time not staring at Praat; Scott Lynn and Emily Morrison have been fantastic hosts.

I also want to thank my family, who have been behind me every step of the way. My parents, Doug Heffner and Peggy Walsh-Heffner, have never been shy about how proud they are of me. No, but really. I once broke a side mirror of my 1994 Buick Regal in high school and my mom told me how proud she was of me. That sort of love and support has been a firm bedrock for the rest of my life. The joy my mom takes from the world around her, and the work ethic of my dad, are both inspirational to me. My brother, Patrick Heffner, has been an inspiring example to live up to… and live *near,* and even live *with* for a brief time during my PhD. I'll finally break the streak of following him around the country, and I'll miss that closeness. However, I know he won't miss me too much given that he has his wonderful wife Patricia Heffner to spend time with, in case he ever needs someone to order the exact same thing that I'd order at a restaurant, or to turn off ESPN because Patrick doesn't need to have it on *all the time*. Patricia, we are so, so, so lucky to have you in the family. And thank you to the extended family, particularly to Garrett Heffner and Joe Walsh for laying the groundwork in the family for another PhD.

And finally, of course, my PhD career would and will forever be impossible think of without Peter Hendrickson. He's put up with a lot over the years—the tendency to randomly travel around the world, to bring up research at the most inopportune times, and to avoid any food with just a hint of spiciness. It's hard to believe it's already been four years together. If he's getting a fraction of what I've gotten from him, I'm glad of it! And to the Hendrickson family (Chris and Kathy Hendrickson, Drew Hendrickson and Diana Schmalzreid, and Tommy Hendrickson and Mary Katica), thanks for having me around so often. And thank you, the reader, for reading this.

Table of Contents

List of Figures

List of Tables

1 Domain-Specificity

"*Stell*? Or *sparklin'*?"

It was the first day of my first conference outside the US, and I found myself across the table from a middle-aged Scottish woman pouring drinks for the conference-goers. I had gone up to her and asked for a glass of water. But her question baffled me. "Stell?" What's "stell"? I had asked her to repeat her request, and again she asked, even more insistently. Well, "sparklin'" was a word that I at least could comprehend, so I asked for that. I was given seltzer water and sent on my way.

As I later realized, I was being asked whether I wanted tap water (i.e., still water) or seltzer water (i.e., sparkling water). My failure to understand the woman likely stemmed from a variety of reasons, not in the least jet lag. In the present dissertation, I consider two aspects of the interaction that likely helped contribute to the miscommunication. I had problems with phonetic *adaptation*, in particular adapting to her Glaswegian dialect, with which I had had very little familiarity before traveling to Scotland. Because of that fact, and the limited speech interaction I had with her before she asked what type of water I would prefer, I had not formed any impression of her or her dialect that would allow me to interpret what she was saying. I also struggled with *categorization*. The [ɪ] in her "still" was lowered enough that I placed it into my [ɛ] category, leading me to hear a non-word ("stell") rather than a word that may have at least given me some hint of what she was asking ("still").

This dissertation probes both categorization and adaptation. In particular, I examine questions of domain-specificity for both tasks. Although the example above comes from the linguistic domain, categorization and adaptation are not unique to language; they are found across a wide variety of domains in various forms. For instance, if you see someone moving at you at a rapid speed, it might be important to know whether they are jogging or running (either at you or away from something). That determination might be a good example of categorization, as you must decide whether the movement you are seeing is an example of a jog or a run. It also provides a good example of adaptation, as it might be that you can judge the person's motion in line with, say, the age of the person, or the person's previous behavior (e.g., have they been maintaining the same pace for a long time?).

For categorization, I look specifically at the acquisition of categories, and whether the processes of category learning used for phonetic categories are the same used to learn other, non-linguistic auditory categories. For adaptation, rather than looking at accent adaptation (which led me astray in Glasgow), I focus on rate adaptation. Just as speakers vary in their accent, they also vary in the rate of speech at which they talk. Taking into account this variation can help to determine, say, where words stop and start in fluent speech. I examine if rate adaptation also helps to explain how viewers determine where events stop and start within action sequences. In doing this, I am exploring questions of domain-specificity; whether and which resources are shared between speech perception and other cognitive abilities.

1.1 What is Domain-Specificity?

Domain-specificity refers to the idea that an aspect of cognition has a particular mode of processing or representation that is unique, unshared with other abilities. As will shortly be seen, defining "aspect of cognition" and "unique" can often be challenging, but the idea has been persistently debated. In the present dissertation, I use "domain" to mean an ecologically-relevant area of knowledge. The “ecologically-relevant” caveat is important; although “parliamentary forms of government” is a domain of knowledge that can be highly relevant to modern society, it is unlikely that knowledge of the fact that Northern Ireland uses a single-transferable vote system for its Assembly elections has ever contributed to anyone’s evolutionary fitness. I first outline general discussions of domain-specificity. I use face perception as an example of a well-trod debate related to domain-specificity. I next review the literature on domain-specificity in speech perception, the object of study in the present dissertation. Finally, I give an overview of the dissertation proper.

1.1.1 Why Be Domain-Specific?

Before diving directly into the idea of domain-specificity in relevant fields, it is important to consider why it might be useful to have domain-specific processes at all. What good might domain-specificity do? Certainly, one answer to that question is “none”. For those who believe in domain-specificity, there are a few reasons given for why it might be useful. They all stem from some version of the adage that “practice makes perfect”. Consider the process of learning to drive a car. At first, every action necessary for driving takes a great deal of effort and a great deal of attention. Yet, over time, many of the skills that are necessary to drive a car become automatic, almost instinctive. Yet it is clear that driving a car is not “ecologically-relevant”; car driving is probably not a domain, at least for the purposes of domain-specific processing. And it is not all that frequent. Although many people drive a car every day, it is usually not for hours at a time, nor incessantly throughout the day. Processing speech, on the other hand, is something that happens often and repeatedly. It is relevant for a great deal of activities, including obtaining sustenance and finding a mate, both of the utmost importance in ecological terms. Processing speech in a domain-specific fashion means devoting a discrete and self-contained chunk of cognitive resources towards speech perception. That chunk is (or becomes) highly specialized, capable of processing *only* speech information. Domain specialization would make the resources devoted to speech very fast at doing their job, as no other tasks would compete to use those resources, but it would also prevent those resources from being used for other tasks if they were lying dormant.

The necessity of domain-specificity and innateness over domain-general learning has been formalized by using three arguments in favor of the idea of innate, domain-specific processes (Cosmides & Tooby, 1994):

1. **Error**: An error in one domain might be beneficial in another. For example, a tendency to consort with kin is beneficial when it comes to altruism (helping kin helps genes that encourage that practice survive), while consorting with kin is not beneficial when selecting a mate (as inbreeding decreases fitness, over time).
2. **Lifetime Incidence**: Many evolutionary advantageous behaviors will not relate to situations observable in all individuals' lifetimes. Many animals have instinctive behavior to flee from forest fires, for example, even though the likelihood that any one animal will experience a forest fire in its lifetime is quite low. This knowledge of the danger of forest cannot be explained using domain-general learning mechanisms, as no single animal

will have enough information in its lifetime to flee, suggesting that innate knowledge may be necessary.

3. **Combinatorial Explosion**: Relying on a domain-general mechanism means that every combination of every cue must be considered when determining a possible response to an environmental trigger. Processing speech becomes impossible if *all* information present for *every* system (auditory, visual, olfactory…) is available to be used as well.

The idea of domain-specificity has often been discussed in tandem with the idea of cognitive modules. According to Fodor (1983), modules are cognitive domains in which information is processed quickly, automatically, and innately by way of a speicifc neural structure. Modules process only certain types of information (domain-specificity), and this information is not subject to interference from other cognitive systems (encapsulation). Fodor argued that only a limited number of cognitive systems (most of those peripheral – e.g., low levels of the visual system) could actually be described as "modular", especially because only those systems could be described as encapsulated.

However, this idea has been extended further in the form of a concept known as massive modularity. For proponents of massive modularity, if a cognitive system takes in certain inputs and produces certain outputs that serve a particular purpose, it makes up a module (Frankenhuis & Ploeger, 2007; Pinker, 2005). For instance, the question of whether any one task is speech-specific becomes a question of whether the *processes* at play within a speech perception task are shared with any other tasks, or whether a single module can accomplish both. According to massive modularity, cognitive modules are present and used at almost every level of human cognition, ranging from the processing of very simple visual patterns to language and moral reasoning (Sperber, 2001). Cognitive modules have been treated as the equivalent of, say, organs, having evolved for a certain, evolutionarily-required purpose (Tooby & Cosmides, 1990).

Ironically, however, one of the fiercest critics of the idea of massive modularity is Fodor himself, as Fodor's claims about what defined a "module" were much more formidable than those usually accepted by proponents of massive modularity. As Carruthers (2005) put it, "it is obvious that by 'module' we can't possibly mean 'Fodor-module', if a thesis of massive mental modularity is to be even remotely plausible" (p. 6). As such, rather than discussing these concepts with relation to speech perception in terms of modularity per se, I focus more specifically on the concept of domain-specificity, something closer in many ways to the principles of massive modularity.

1.1.2 Outside Language: Face Perception

One area of study in which there has been vigorous debate about domain-specificity is in the perception of faces. Faces seem to be processed in a "special" way. For example, faces are processed holistically (as a unified chunk) rather than in a piecemeal fashion as other visual objects are (Farah, Wilson, Drain, & Tanaka, 1998). Humans are drawn to look at other human faces; in one study, even neonates as young as 9 minutes old seem to prefer looking at faces over artificially-scrambled face-like images (Goren, Sarty, & Wu, 1975). The developmental time course of the preference for faces indicates that 6-month-olds can distinguish both individual human faces and individual monkey faces; however, by 9 months of age, infants can no longer distinguish monkey faces, while preserving an adult-like ability to distinguish human ones (Pascalis, de Haan, & Nelson, 2002), although a 6-month-old-like ability to distinguish monkey faces can persist with an appropriate training regimen (L. S. Scott & Monesson, 2009). A failure to attend (or properly attend) to faces is often seen as a sign of clinical disorder, as in the case of people with an autism spectrum disorder, who do not show typical adults' strong preference for looking at the eyes of actors in short film clips (Klin, Jones, Schultz, Volkmar, & Cohen, 2002).

Face perception is intimately linked with the fusiform face area (FFA), a region in the right fusiform gyrus that shows persistent activation across a wide variety of face-perception-linked tasks. Activation in the FFA has been correlated, for example, with the perception of the face-vase illusion (see Figure 1, below), where the percept of the image switches back and forth between a vase (in white) or two faces gazing at each other (in black). Greater FFA activation in one study was linked to periods in which the bistable illusion is seen as two faces rather than a single vase (Andrews, Schluppeck, Homfray, Matthews, & Blakemore, 2002). Even infants as young as 4 to 6 months old show face-selective regions in visual cortex, as assessed using fMRI study (Deen et al., 2017).

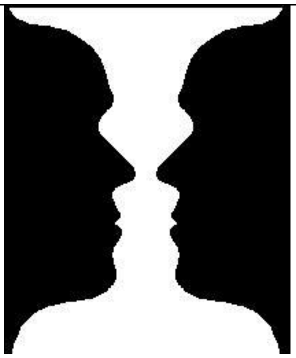


Figure 1. The face-vase illusion

Is face processing domain-specific, then? It is certainly "special", but none of the evidence cited above necessarily requires domain-specific processing. After all, faces are visual objects that humans encounter on nearly a daily basis and that a serve special importance, making it possible to imagine that the effects arise as a result of training. However, it is also true that the fact that there is a developmental trajectory to face perception does not rule out the idea that face perception is domain-specific.

The side in favor of viewing face perception as domain-specific has largely been led by Nancy Kanwisher (Kanwisher, 2000). Behaviorally, as was alluded to briefly, there seems to be something special and holistic about face processing (Farah et al., 1998). This can be seen fairly straightforwardly when comparing the processing of faces that are upside-down, which are very hard to tell apart, from faces that are right-side-up, which are quite easy to distinguish (Yin, 1969), a fact that is quite different from the vast majority of other visual objects. This suggests that right-side-up faces are processed in a special way.

Brain regions such as the FFA also provide evidence for domain-specificity. Functional regions in the FFA that respond more generally to faces over other objects also showed greater responses to faces on a battery of follow-up tasks, an idea that was argued to provide strong evidence for domain-specificity, with the FFA being the “hub” of face processing (Kanwisher, McDermott, & Chun, 1997). Similar gradients were not observed for participants having to distinguish between other classes of objects, such as guitars, birds, flowers, or cars (Grill-Spector, Knouf, & Kanwisher, 2004); subtle manipulations of the visual properties of houses, for instance, do not lead to fMRI activation patterns that come at all close to those seen for faces (Yovel & Kanwisher, 2004).

A third source of evidence comes from the neuropsychological literature. Prosopagnosia, or face-blindness, refers to an inability to distinguish individual faces. It is strongly associated with lesions or malformations of the FFA (De Renzi, 1986). Prosopagnosia can be quite specific to human faces; indeed, one patient with prosopagnosia had no difficulties whatsoever after starting a new life as a farmer in recognizing and distinguishing sheep, despite difficulties in recognizing and distinguishing human faces (McNeil & Warrington, 1993). This can be compared to patients with lesions elsewhere who have no problem distinguishing faces but have a hard time recognizing other objects or visual words (Moscovitch, Winocur, & Behrmann, 1997). These results suggest the presence of a double dissociation, as decrements in face recognition do not necessarily imply concomitant decreases in object recognition (and vice-versa).

The main critics of the idea of domain-specificity in face processing have proposed instead that face perception is underlain by domain-general mechanisms responsible for expertise (Bukach, Gauthier, & Tarr, 2006). The behavioral results that are said to suggest that faces are processed in a "special" way, and the privileged status of the FFA, are both said to derive from expertise. The FFA, rather than being face-specific, is instead seen as an "expertise area". Much of the evidence for domain-general processing comes from studies that have used novel objects referred to as greebles to probe the question. Greebles are invented visual objects that share some of the complexity of faces, but are not readily perceived as face-like. Novice greeble-watchers do not show any face-like perceptual effects when processing faces; for example, shuffling the different components of greebles does not affect their discrimination abilities. However, given enough training (between 7 and 10 hours), now-expert greeble-watchers show patterns of discrimination that resemble those found in face perception studies (Gauthier & Tarr, 1997). This is sometimes accompanied by activation of the FFA (Tarr & Gauthier, 2000). People with an autism spectrum disorder show evidence of deficits in their ability to discriminate greebles that resemble those found in their perception of faces (Scherf, Behrmann, Minshew, & Luna, 2008). The FFA also lights up when experts are discriminating between the objects of their expertise; for example, when bird-watchers are discriminating between individual birds, or when car experts are discriminating between individual cars (Gauthier, Skudlarski, Gore, & Anderson, 2000).

Yet not all is rosy for the idea that face perception is domain-general. The behavioral evidence, for example, is mixed at best; neither greebles nor cars (to car experts) nor cells (to cell biologists) show evidence for being processed in a holistic manner as strongly as faces are (McKone, Kanwisher, & Duchaine, 2007). Prosopagnosia seems to spare greeble learning, suggesting that damage to the brain areas that underpin face perception spares the acquisition of greeble expertise (Duchaine, Dingle, Butterworth, & Nakayama, 2004). This has led some to propose a middle ground between the theories outlined above: the FFA starts with some pre-existing biases in favor of becoming face-specific, then, with sufficient experience over a number of years, gains the expertise necessary be considered a “domain-specific” area (Cohen Kadosh & Johnson, 2007). I question the extent to which this is truly a middle ground, however. The full development of a domain-specific face processing ability could also rely on the (reasonable) expectation that enough experience be provided to result in its final form; in this case, it seems quite reasonable that (non-visually-impaired) adults would have plenty of experience with faces to allow face-specific areas to develop. Overall, I find the evidence in favor of domain-specificity in the face perception system to be fairly convincing.

1.2 Domain-Specificity in Speech Perception

The discussion of domain-specificity across cognitive domains has always been strongly linked to the study of language. It is Chomsky's initial explorations of language that are often credited as the first evidence for the necessity of a domain-specific learning mechanism in language, as Chomsky gave evidence that syntactic processing and acquisition demand mechanisms particular to the domain of language (L. A. Hirschfeld & Gelman, 1994). Although most of Chomsky's speculation about the domain-specific nature of language relate to syntactic processing, not the questions of speech perception largely focused on in the present project, there is no question that language is a common touchstone in this literature. In the words of Alvin Liberman, who was probably the person most associated with a domain-specific view of speech understanding, the idea is that speech perception is "a distinctively phonetic process specifically adapted to the unique characteristics of the speech code" (Liberman, 1982, p. 152). In the section below, I outline the literature related to this idea, including both arguments for and arguments against the idea of domain-specificity in speech perception, often cast in terms of the hypothesis that "speech is special".

1.2.1 Evidence for Domain-Specificity in Speech Perception

The primary source of arguments in favor of the idea of domain-specificity in speech perception come from proponents of motor theories of speech perception. Motor theories have long been associated with evidence in favor of speech-specific processes in auditory perception. However, evidence is not limited to a single theoretical perspective, as important as that perspective has been. Other evidence comes from neuroimaging, neuropsychology, and developmental findings. These are discussed below.

1.2.1.1 Motor Theories of Speech Perception

Motor theories of speech perception generally have taken a very strong view that speech is special. Proponents of motor theories were among the first to appropriate the idea of a cognitive module in the sense suggested by Fodor to a non-peripheral cognitive domain. Motor theories tend to eschew the acoustic signal as the fundamental correlate of categories, and to treat motor gestures as the underlying "fundamental". Instead of positing invariant *acoustic* features related to speech sounds, proponents of motor theories of speech perception instead favor invariant *motor* features, which have variable (but still predictable) acoustic correlates (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). Speech is thereby special because the nature of the speech signal, as well as the complex relationships between production and perception that motor theorists claim are necessary to comprehend that signal, require speech-specific mechanisms (Liberman, 1982). To comprehend speech using only principles of "a generally auditory sort" was described as "hardly conceivable" (Liberman, 1982, p. 151).

How did such a view arise? Very early research at Haskins Laboratories intended to create an auditory "alphabet" for use by the blind, comprised of an arbitrary sequence of non-linguistic sounds. However, when played back for the people who the alphabet was designed to benefit, the alphabet was incomprehensible, despite the fact that it was being played at a speech-like tempo (Galantucci, Fowler, & Turvey, 2006). This led the researchers at Haskins Labs to investigate the speech signal in greater detail. What they found was that the speech signal is not like an alphabet at all; the acoustic features of the sounds that speakers produce are variable. In particular, sounds are coarticulated, with a particular speech sound (say, [p]) being produced differently depending on the adjacent sounds (Liberman, 1957; Liberman, Delattre, & Cooper, 1952). Speakers anticipate the sound yet to come, while still showing delayed effects of the sounds recently produced. The modern understanding is that speech is highly redundant, with multiple overlapping cues to individual sounds that might occur several syllables away from where the sound is perceived. For example, the distinction between [i] and [a] can be cued by vocalic differences multiple syllables in advance of that distinction (Grosvald, 2009).

Such findings on their own, of course, would not necessarily demand a motor theory of speech perception. After all, if coarticulation was accompanied by entirely *predictable* acoustic cues, listeners could simply learn the distributed acoustic signatures that correspond to each sound. However, motor theorists have argued, the acoustic cues that accompany speech sounds are unpredictable without recourse to knowledge of the articulatory patterns of speech (Liberman, 1957; Liberman et al., 1967). For example, the frequency transitions at the onset of vowels following [d] involve increases for front vowels (such as [i]) and decreases for back vowels (such as [u]), such that the formant transitions appear to "point at" a frequency of approximately 1800 Hz; however, this transition must not be complete (that is, it must not "touch" 1800 Hz) for the sound to be perceived as [d], which motor theorists believed reflected the motor gestures necessary in order to produce [d] (Delattre, Liberman, & Cooper, 1955).

These findings inform the idea of speech's special properties. Over time, speech perception came to be seen as the outcome of speech-specific processing that is responsible for linking acoustic properties of the speech signal with invariant and potentially innate articulatory specifications (Liberman & Mattingly, 1985, 1989). In other words, speech perception is modular. “Module” to Liberman was defined in Fodorian terms (Fodor, 1983), and is comparable to, say, bats’ specialized echolocation abilities. The speech module was said to stand separately from non-speech-specific acoustic processing of factors such as pitch, timbre, and volume. It was instead supposed to carve out and separate the information needed to understand speech from other information (Liberman & Mattingly, 1989).

A phenomenon known as “duplex perception” provided evidence given to support the idea of speech-specificity. In duplex perception paradigms, stops (or, to be more precise, the vowel transitions corresponding to different voiced stops) are presented binaurally; in one ear, listeners hear the first two formants, whereas, in a third, listeners hear the F3 transition (Mann & Liberman, 1983; Repp, Milburn, & Ashkenas, 1983). What percept results depends on the intensity of the F3 transition; if it is soft enough, only the a stop consonant is heard, while a louder F3 leads to the simultaneous perception of a stop and a whistle corresponding to the frequency of the F3 (Whalen & Liberman, 1987). The claim is that the fact that, because the F3 transition can be perceived either as belonging to the speech sound or as a “chirp” on its own (or both, simultaneously), it is being evaluated by two separate perceptual systems.

Sine-wave speech provides another line of support for the idea that speech is special, although its implications are also somewhat troubling to motor theories. Sine-wave speech, as its name implies, is created by combining sinusoidal signals that approximate formant frequencies. Without any cuing, sine-wave speech does not sound much like speech; it might sound a bit like whistling, or a “science fiction sound”.

Despite this, it has many of the time- and frequency-varying properties of speech, and given an appropriate prompt, listeners can successfully perceive words that are a part of the signal (Remez, Rubin, Pisoni, & Carrell, 1981). Listeners can perceive and adjust to variation between talkers when listening to sine-wave speech, similar to the adjustments that are present for normal speech (Remez, Rubin, Nygaard, & Howell, 1987; Sheffert, Pisoni, Fellowes, & Remez, 2002). The perception of sine-wave speech can be modulated on a trial-by-trial basis; it has been argued that listeners can both attend to sine-wave speech and attend to the individual sinusoids in the signal, depending on task demands (Remez, Pardo, Piorkowski, & Rubin, 2001). These results generally indicate that, for sufficiently ambiguous signals, the processing of the signal as speech can be switched on and off, in line with the idea of speech-specific processing. However, these findings are also challenging to accommodate within motor theoretic accounts of speech perception, as the impoverished nature of the sine-wave signal make articulatory gestures more challenging to perceive. Indeed, these findings were used to advance a view that was auditory-based and domain-specific (Remez, 1989; Remez, Rubin, Berns, Pardo, & Lang, 1994).

Motor theories came under increased skepticism in the 1990s and early 2000s, such that a review from 2006 by some of the most prominent contemporary motor theorists acknowledged that it had “few proponents within the field of speech perception” and that “many authors cite it primarily to offer critical commentary” (Galantucci et al., 2006, p. 361). Take duplex perception as an example. Fowler and Rosenblum (1990) compared the duplex perception of F3 transitions in stops with the duplex perception of a slamming door. Just as with speech sounds, the slamming door sound was split into two parts, segregated by frequency, and played simultaneously with each part played to a different ear. Listeners were taught to label the unmodified door-slamming sound as a “metal door”, the low-frequency portion of the door-slamming sound was a “wooden door”, and the high-frequency portion of the door-slamming sound as a “shaking sound”. The patterns observed often resembled those seen in speech (i.e., that the perception of the high-frequency portion of the signal was modulated by its intensity, but often fused with the percept in the other ear). As it is highly unlikely that the sound of a door slamming would be processed by a speech perception system, a likelier explanation was that the processes that caused duplex perception in speech were identical to those used to process the door slamming sounds, and that, therefore, duplex perception cannot be used to justify speech specificity.

Some authors have brought the motor perspective on speech perception further by arguing for a “direct realist” view of perception, where the perception of speech is said to directly arise out of the perception of motor gestures, with only the minimal possible recourse to the acoustic information being received. This is affected by direct motor activation of speech sound production during the process of perception (Best, 1995; C. A. Fowler, 1986). Interestingly, these theories, though directly building on motor theories, generally make the opposite claim as motor theories on whether speech is special. Instead, they claim that a very wide variety of sounds, not just speech, are processed through the use of the motor system. Thus, speech is not special, and it is not special specifically because it, like most other perceptual abilities, is actually undergirded by systems of production (Willems & Hagoort, 2007; Worgan & Moore, 2010). These theories have seen an upswing of late as a result of an interest in a class of neurons called "mirror neurons", which are said to have properties reminiscent of the direct realist approaches that coupled perception directly to production (Schwartz, Basirat, Ménard, & Sato, 2012). Although the contributions of mirror neurons and the production system more generally to speech perception have been sharply questioned (Hickok, 2009; Lotto, Hickok, & Holt, 2009; S. K. Scott, McGettigan, & Eisner, 2009), there is no question that such findings have led to a resurgence in interest in motor theories of speech perception. Yet that interest is divorced from one of motor theory's main tenants: speech is special. Speech perception is a part of a gestural perception that spans many domains.

1.2.1.2 Neural Architecture for Speech Perception

The neural architecture of speech perception is also brought in to support the idea of speech-specificity in speech perception. The uncovering of speech-specialized brain regions has become somewhat of a cottage industry; many studies have uncovered evidence for "speech regions" in the brain (Price, 2012). Depending on the study, neural activation to speech has been compared to, say, noise, tones, and reversed speech (Binder et al., 2000), to laughter and other environmental sounds (Meyer, Zysset, von Cramon, & Alter, 2005), and to stimuli parametrically varied according to several potentially relevant dimensions (Benson et al., 2001; Leaver & Rauschecker, 2010). The regions uncovered have clustered in left temporal regions and left pars opercularis, a region that in its many forms (Broca's Area, BA 44/45, left inferior frontal gyrus) has been repeatedly cited in neural studies of language (Price, 2012). There is also a broad consensus that speech is processed along two separate streams, one ventral and one dorsal (Hickok & Poeppel, 2007; S. K. Scott, Blank, Rosen, & Wise, 2000), a fact that has been tied into similar pathways in auditory and visual perception in the animal literature (Petkov, Logothetis, & Obleser, 2009; Rauschecker & Scott, 2009).

The uncovering of these speech-sensitive regions, however, has been greeted with some skepticism to the idea that these are relevant to the idea of speech being special (Price, Thierry, & Griffiths, 2005). Consider the comparisons that are being run to establish whether the brain regions are activated: the hemodynamic response to speech is compared to the hemodynamic response to non-speech stimuli, whether cough or instrumental music or noise. This means that any activation is by its nature *relative*, implying that the same regions might also be involved in processing the non-speech signals, but that they are simply less activated in one condition than another. Functional MRI is by its nature spatially inexact. Patterns of activation show brain regions measured in terms of cubic millimeters, which each contain many thousands of neurons. Instead of showing speech specificity, the fMRI findings might instead reflect differential demands on pre-existing auditory resources.

Better evidence for speech-specificity in the brain comes from neuropsychological studies. Two conditions, pure word deafness and auditory agnosia, suggest the presence of the holy grail of neuropsychological evidence: the vaunted double dissociation. A double dissociation arises when damage to one part of the brain (generally through stroke) leads to impaired functioning for a first behavior but not a second, while damage to a different part of the brain leads to impaired functioning for the second but not the first. The idea is that the two behaviors *cannot* share common resources (or, at least, there must be some resources used by each process that are not used by the other) if there are some parts of the brain that selectively respond to each function independently.

For speech perception and auditory perception, the two neuropsychological disorders that doubly dissociate are pure word deafness and auditory agnosia (Poeppel, 2001). Pure word deafness refers to a neuropsychological condition in which patients with brain damage (generally in temporal cortex) are incapable of understanding speech, while maintaining all other auditory processing abilities (Auerbach, Allard, Naeser, Alexander, & Albert, 1982; E. M. Saffran, Marin, & Yeni-Komshian, 1976; Tanaka, Yamadori, & Mori, 1987). It can be doubly-dissociated from auditory agnosia, where patients are generally unable to understand or recognize non-speech environmental sounds (e.g., keys jangling) but are able to comprehend speech without any problem (Fujii et al., 1990; Lambert, Eustache, Lechevalier, Rossa, & Viader, 1989; Spreen, Benton, & Fincham, 1965). Auditory agnosias often progress from more severe conditions (Taniwaki, Tagawa, Sato, & Iino, 2000), and are accompanied by neural plasticity, such as changes in hemispheric specialization for language (Saygin, Leech, & Dick, 2010). The presence of this double dissociation suggests that at least some of the neural resources necessary for speech are different from those used in other forms of auditory processing (but see Pinard, Chertkow, Black, & Peretz, 2002 for an alternative perspective on pure word deafness).

1.2.1.3 Developmental Evidence for Domain-Specificity

Developmental findings have also been used to justify the idea of speech-specific processing of speech information. The idea is that if young enough infants show an ability to attend to and process speech information, this could indicate an innate proclivity for speech perception that is unmatched by other types of auditory processing. One early study used a preferential looking paradigm to compare 4.5-month-old infants' looking times to speech versus those to white noise and found that the 4.5-month-olds were more likely to look at a visual object linked to the speech sound than one corresponding to the white noise (Colombo & Bundy, 1981). However, white noise makes a poor comparison to speech, as it lacks much of the acoustic complexity of speech. Athena Vouloumanos and colleagues have recently rectified many of the issues identified in that and other studies of infants' speech preferences. Using a similar procedure, Vouloumanos and Werker (2004) found that infants as young as 2 months old preferred listening to natural speech more than listening to sine-wave speech-like auditory signals. This preference was later extended to neonates (just 1 to 4 days old) using a high-amplitude sucking paradigm (Vouloumanos & Werker, 2007). Later experiments along these lines showed that infants at young as 6 months recognize that speech involves the communication of information (Vouloumanos, Martin, & Onishi, 2014), and that the strength of this bias at 12 months of age can predict language outcomes at 18 months (Vouloumanos & Curtin, 2014), showing that the speech bias has practical implications as well.

It is not just the case that infants are drawn to speech at a young age; they also show evidence for processing it in a way that sometimes resembles adults. For example, they are capable of processing subphonemic detail in a way that resembles adult listeners. 2-month-olds can differentiate variants (or allophones) of stop and liquid sounds that can be used to signal word boundaries (Hohne & Jusczyk, 1994). 3.5-month-olds apparently compensate for coarticulation and cue weighting in the perception of consonant voicing and manner contrasts (J. L. Miller & Eimas, 1983). 3-month-olds can integrate information presented across two ears in order to perceive a sound as a [da] or a [ga], a fact that resembles the previous studies of duplex perception (Eimas & Miller, 1992).

Infants also show perceptual constraints that resemble those present for adult speakers. For example, consider the ability of 6-to-8-month-old infants being raised in an English-speaking environment to perceive categories of stops within a place continuum ranging from a bilabial to a retroflex voiced stop. The infants could distinguish pairs of consonants that are treated as different categories in both Hindi and English (e.g., [b] and [d]) and Hindi alone (e.g., [d] and [ɖ], which are both treated as examples of [d] for English speakers), but could not differentiate pairs of consonants that are indistinguishable to speakers of both languages, such as tokens that would both be considered examples of [d] in English and [ɖ] in Hindi (Werker & Lalonde, 1988). Infants as young as 4.5 months show constraints on phonetic perception that resemble faithfulness and markedness constraints in Optimality Theory (Prince & Smolensky, 2004), indicating a bias towards simpler and less variable sound sequences (Jusczyk, Smolensky, & Allocco, 2002). In one particularly fascinating study, 6-month-old infants were unable to distinguish dental and retroflex stops (such as those used in Hindi) when they were given a flat teether that impaired their ability to control the movement of their tongue (Bruderer, Danielson, Kandhadai, & Werker, 2015), which contrasts with their ability to successfully discriminate when given a gummy teether that does not affect tongue movements. This suggests that some degree of sound-to-motor correspondence is known by infants as young as 6 months old, as proper control of the tongue is necessary to differentially produce the dental and retroflex stops.

1.2.2 Evidence against Domain-Specificity

Motor theories have been criticized since virtually their inception (Lane, 1965). For example, opponents of motor theories cite the complexity of routing through a separate motor representation rather than more directly focusing on the sound signal that is supposed to transmit articulatory detail. In direct contrast to motor theories of speech perception, general auditory theories take almost the polar opposite view of phonetic perception. Under general auditory theories, speech is, in fact, *not* a mode of processing that requires special abilities. Instead, general auditory theories treat speech sound perception as a special instance of auditory processing. Experiments that test general auditory theories often involve creating non-linguistic analogues to typical speech perception tasks, with the idea that the same pressures that are exerted on speech sound perception are also exerted on those analogues. Many of the supposed linguistic effects from previous studies are found in other auditory domains (Diehl, Lotto, & Holt, 2004; Holt & Lotto, 2008). Evidence for general auditory views generally comes from three sources: context effects on phonetic categorization, animal models of speech perception, and the role of expertise in phonetic categorization.

1.2.2.1 Cross-Domain Context Effects

The first line of evidence in favor of general auditory models takes some of the primary sources of evidence cited in favor of motor theories of speech perception and turns it on its head: coarticulation (Holt & Kluender, 2000). Even without evidence from other sources, there was some speculation that coarticulatory effects could best be explained through acoustic models, not ones that depended on the gestural underpinnings that caused them (Massaro & Oden, 1980). For example, the perception of the middle vowel in CVC sequences is influenced much more strongly by coarticulation if the consonants within the sequence are both voiced than if they are both voiceless. This arises even though the consonants involved require (roughly) analogous gestures; for instance, both [t] and [d] involve stopping the airflow entirely using the tip of the tongue. Under motor theories, this is puzzling; the same motor gestures should lead to similar coarticulatory effects. Proponents of general auditory theories suggest instead that the differences between the voiceless and voiced consonants' coarticulatory effects instead arise from the acoustic information that characterizes the consonants (Holt, Lotto, & Kluender, 2000). However, the primary mode of attack against the idea of coarticulatory effects showing speech-specificity was along the lines of the same one used to challenge duplex perception (C. A. Fowler & Rosenblum, 1990); namely, by finding a non-speech analogue that could trigger effects that resembled those in speech.

One case study relates to consonantal sequences of a liquid (such as [l] or [ɹ]) followed by a voiced stop (ambiguous between [g] or [d]). Both English and Japanese listeners alike are more likely to perceive the ambiguous stop as a [g] when the preceding syllable ends with [l] than when the preceding syllable ends with [ɹ] (Mann, 1986), a fact that is particularly remarkable because Japanese speakers cannot perceptually distinguish [l] and [ɹ]. For motor theories of speech perception, this results from the differing tongue gestures used to produce [l] and [ɹ], which have lawful relationships with the F3 values that typically signal the difference between [g] and [d]. Indeed, this was first treated as an excellent test case for motor theories. However, under general auditory theories, such findings can be explained in terms of frequency contrast; the F3 frequency information that distinguishes [l] from [ɹ] might also lead to differences in the perception of the F3 frequency information that distinguishes [g] from [d].

To test the frequency contrast account, proponents of general auditory theories have used pure tones to modulate phonetic boundaries. Pure tones are often used for experiments along these lines because it is clear that they could not be seen as a part of any purely phonetic module. Thus, the influence of pure tones on phonetic categories is taken as evidence that information outside the linguistic signal can influence speech perception. Indeed, non-linguistic tones at frequencies at or around those characteristic of the F3 of [l] and [ɹ] also lead to “coarticulatory” effects on following stops; an [l]-like pitch leads to more [g] reports (Lotto & Kluender, 1998). Long-term distributions of pure tones around [ɹ]- or [l]-like frequencies can also trigger similar coarticulation-like effects (Holt, 2005, 2006), regardless of the pitch immediately preceding the stop. Similar contingencies have been found for the influence of the rate of non-linguistic tones on tokens ambiguous between [b] and [w] (Wade & Holt, 2005).

Of course, this evidence has not gone unchallenged. One of the most interesting pieces of evidence against this comes from Tamil, a Dravidian language spoken in South India. Tamil has two liquids, [ɾ] and [ɭ], which are both perceived by English speakers as examples of [ɹ]. Acoustically, both resemble [ɹ] in their third formant frequencies. In terms of the articulatory underpinnings of each sound, [ɾ] is produced with the tongue close to the front of the mouth, while [ɭ] is produced with a retroflex tongue tip, reversing the pattern found for English’s [ɹ] and [l]. According to motor theories of language, the difference in tongue gestures should lead to a reversal of the pattern found in English, with more [g] responses after [r] rather than after [ɭ]. According to general auditory theories, meanwhile, [g] responses should be approximately equal across the two conditions, as the F3 frequencies are the same across each liquid. Viswanathan, Magnuson, and Fowler (2010) found that the results much more closely tracked the predictions of motor theories of speech perception. Strikingly, these effects were not found for any number of non-speech analogues of the Tamil sounds; non-speech tones with similar frequency information tended to lead listeners to report [g] at a rate similar to that of [ɹ], resembling the acoustic qualities of [ɹ], [ɾ], and [ɭ].

1.2.2.2 Animal Models of Speech Perception

Another source of evidence cited in favor of general auditory theories of speech perception comes from the non-human animal literature. If, as motor theories argue, speech perception is a uniquely human capability, born of innate knowledge of mappings between sounds and articulations, non-human animals (who would have no reason to have knowledge of human articulators) should not be able to perceive speech in a human-like fashion. At a broad level, it may be that such criticisms miss the point; it is rarely the case that the non-human animal studies of speech perception propose, say, a common mechanism to explain human and non-human animal perception of speech, such that the ability to perceive speech in a human-like fashion represents a homologous development (Trout, 2001). Still, non-human animal studies of speech perception are ubiquitous in this literature. Particularly important models have included the chinchilla (*Chinchilla lanigera*), the Japanese quail (*Coturnix japonica*), the budgerigar (*Melopsittacus undulatus*), and non-human primates (including chimpanzees and macaques).

Chinchillas were some of the first animals to have their perception of human speech studied. Their hearing acuity across frequencies is broadly similar to those of humans, making them an excellent model animal to study aspects of human hearing (J. D. Miller, 1970). Kuhl and Miller (1975) trained chinchillas to categorize stop continua that varied in voice onset time (VOT), with intermediate steps being heard as either a [t] or a [d], by using an avoidance condition procedure. Under that procedure, the chinchillas were taught to rush to the opposite side of their cage from a water tube when they heard one of the endpoints on the continuum, or else a shock would be delivered. After being trained to distinguish the [t] and [d] endpoints, the intermediate steps were played for the chinchillas. The proportion of trials in which the chinchillas fled the water tube was treated as an index of how often they heard that sound as belonging to the trained category. Chinchillas showed strikingly similar response patterns to humans asked to identify the exact same items along a continuum; as with humans, chinchillas categorically perceived the items, with stimuli having a VOT of less than 20ms being reliably categorized as a [d], stimuli having a VOT of more than 40ms being reliably categorized as a [t], and a sharp identification gradient intermediate between those values. Follow-up results indicated that the training could also extend to other stop categories, with training on [t] and [d] extending to both bilabial and velar stop categories, again in a way that strongly resembled native English speakers' perception of the stops (Kuhl & Miller, 1978). Recordings from chinchilla auditory nerves indicated that these findings may have stemmed in part from aspects of auditory nerve responses in response to auditory discontinuities, with a natural discontinuity in neural responses roughly aligning with the discontinuity present in perceptual responses (Sinex, McDonald, & Mott, 1991). Chinchillas also seemed to ignore phonetic detail that is used for speaker identification when distinguishing between [a] and [i] (Burdick & Miller, 1975).

Human-like performance has also been observed in avian models. Songbird communication provides some of the best analogues to human vocal communication (Soha & Peters, 2015) as songbirds demonstrate evidence of vocal learning, the ability to learn and imitate non-innate vocalizations. Songbirds are in fact one of the very few classes of animals that have this ability, along with bats, cetaceans (such as dolphins), parrots, and hummingbirds (Jarvis, 2004) and elephants (Poole, Tyack, Stoeger-Horwath, & Watwood, 2005). Vocal learning is, of course, one of the hallmarks of human linguistic competence: humans are capable of learning new words, and even new languages, throughout the lifespan. This makes songbird communication an alluring target for speech perception studies.

Two of the most-studied songbird species with regard to speech perception are Japanese quail and budgerigars. Like humans, budgerigars are affected by the rate of context syllables in distinguishing [b] and [w] (Dent, Brittan-Powell, Dooling, & Pierce, 1997). They parse VOT continua in a categorical fashion, with stimulus labeling roughly matching that for humans (Dooling, Okanoya, & Brown, 1989). And they (but not zebra finches, another common bird model) use F3 to distinguish [ɹ] and [l] in categorical fashion (Dooling, Best, & Brown, 1995). Studies of Japanese quail went even further by exploring the coarticulatory effects explored in detail above. Japanese quail trained to peck at different keys when they heard [g] and [d] sounds were more likely to peck at the [g] key when an ambiguous stop was preceded by [l] than when it was preceded by [ɹ] (Lotto, Kluender, & Holt, 1997). These results suggest that it is not just humans that show human-like categorical effects in perceiving speech sounds.

Interestingly, the speech perception capabilities of our close animal relatives, primates, remains relatively unexplored. This may be in part because other primates are not vocal learners (Egnor & Hauser, 2004). However, chimpanzee vocal tracts have very recently been described as capable of producing sounds that closely resemble human speech sounds (Fitch, de Boer, Mathur, & Ghazanfar, 2016), and the neural organization of at least primary auditory cortex when listening to speech appears to be similar to humans (Steinschneider, Nourski, & Fishman, 2013). Studies of speech perception in non-human primates have been mixed in their findings. Although macaques do not show similar patterns of categorization as English-speaking humans in distinguishing [ɹ] and [l], putting the boundary between each liquid category in a different location, they do combine cues to the liquid distinction in a way that approximates human abilities (Sinnott & Brown, 1997). Those cue integration relationships do not hold, however, when considering differential cues to the perception of *say* versus *stay* (Sinnott & Saporita, 2000). Macaques are also worse than humans when extending their knowledge of trained sound categories (such as [b] and [d]) to new categories (Sinnott & Williamson, 1999), and do not integrate formant frequencies to categorize vowel sounds in a similar way to humans (Sinnott, Brown, Malik, & Kressley, 1997).

1.2.2.3 Expertise

A final source of evidence given in favor of general auditory theories of speech perception relates to the idea of category expertise in phonetic perception. The idea is that, to the extent that people perceive speech sounds differently from other types of auditory perception, these differences follow mostly from the massive amounts of experience that humans have in hearing language, not from any abilities special to language per se (Lotto, 2000). This tracks closely many of the debates on innateness versus experience in face perception; general auditory theories take the same position as proponents of domain-general approaches to face perception (Tarr & Gauthier, 2000) in saying that erstwhile speech specialization is solely the result of expertise. Some aspects of the categories that are learned may obey some naturally-occurring constraints in acoustic perception (Holt, Lotto, & Diehl, 2004), but even aspects of those constraints may require learning (Holt, Lotto, & Kluender, 2001).

This has led proponents of general auditory theories to link speech sound categorization to questions of categorization in cognitive psychology more generally (Holt & Lotto, 2010), with the suggestion that learning processes within language should resemble those outside of language (Liu & Holt, 2011). Interesting enough, these theories would also predict that language categorization might bleed back over into non-linguistic tones. There is some evidence that English and Japanese speakers differ in their ability to learn [ɹ]- and [l]-like non-speech sounds (Iverson, Wagner, & Rosen, 2016). Given the focus of the dissertation, the importance of learning and expertise in categorization will be returned to in much more detail in Section 2.

1.2.3 Summary

The results of the experiments summarized above leave the field of speech perception at an interesting point. Despite rollicking debates historically, many of the theoretical perspectives with regard to speech perception seem to have converged on the idea that speech is *not* particularly special. That is, speech perception is not separate from auditory perception. The question debated is whether the fundamental units of speech perception are acoustic (Holt & Lotto, 2008) or gestural (Galantucci et al., 2006) in nature; but there are not many advocates of the idea that, whatever those units may be, they are perceived or processed in a way that is different from other entities in the world. Yet speech must *somehow* be different from the rest; after all, speech is used as an input to broader language systems, such as syntax and semantics. Neither the sound of jangling keys nor the motor gestures associated with jangling keys can become a part of a syntactic phrase (Poeppel, Idsardi, & van Wassenhove, 2008). And the studies of infant speech perception, do convincingly paint a picture that infants can distinguish and attend to speech information to the exclusion of other, similar acoustic signals (Vouloumanos & Werker, 2007). So are there any principles that are domain-specific in the processing of speech? To the extent that there are, what are they, and where did they come from? To the extent that abilities are shared, what elements are shared?

1.3 The Present Dissertation

The present project seeks to explore those questions in two domains: categorization and adaptation. I hope to refine whether phonetic learning diverges from category learning more generally, as well as whether the principles that allow listeners in speech to adapt to speakers of different rates are also present when perceiving visual actions. In doing so, this research will help to establish the extent to which speech is special. The projects related to category learning below follow fairly conventional lines, bringing thoughts and ideas from the domain-general category-learning literature to bear on phonetic categories. The projects related to adaptation, however, take a different tack; they ask instead to what extent the principles of phonetic adaptation and word segmentation can be applied to the visual event segmentation literature, which represents a novel directionality for most studies examining the domain-generality of phonetic processing. Given the split nature of the projects that make up this dissertation, it is split into two main sections.

Section 2 centers on category learning. Chapter 2.1 gives an overview of the category-learning literature, focusing on theories of category learning both outside and inside phonetics. Chapter 2.2 reports the results of previous experiments probing phonetic category learning that I conducted while at the University of Maryland. I uncovered a bias in phonetic category learning against the easy acquisition of disjunctive categories (i.e., categories that "skip around" in phonetic space), which does not much resemble proposals outside the speech literature. Chapter 2.3 takes the methodologies I used to examine language and applies them to study the acquisition of non-linguistic categories, in particular categories of musical instruments. The biases present in the phonetic learning experiment are not present in the acquisition of musical instrument categories, suggesting that some aspects of phonetic category learning are domain-specific.

Section 3 focuses on rate adaptation in event segmentation. In Chapter 3.1, I discuss the concept of adaptation, especially rate adaptation, in phonetics, as well as its consequences for word segmentation. Chapter 3.2 includes a previous experiment of mine exploring rate adaptation in Arabic, which showed strong rate adaptation effects. Chapter 3.3 takes the rate adaptation literature within phonetics and applies the methodologies used there to the segmentation of visual events. I find evidence for rate adaptation in event perception that I believe is the first of its kind and has striking similarities to rate adaptation in speech perception.

A final chapter, Chapter 4, provides conclusions and final discussion. I advance the idea that both phonetic learning and phonetic adaptation might themselves form a single processing domain that requires “phonetic plasticity” that can help in both learning new speech sound categories and adapting to variation in old ones. I talk about possible future studies of this idea, and other extensions to the present experiments. Particular attention is paid to possible neural underpinnings of phonetic plasticity, as well as applications of that idea to disordered populations.

2 Category Learning

2.1 Background

What makes a mammal a mammal? Is it something about fur? Live birth? Milk production? Ear bones? A certain type of jaw? Modern scientific theories propose some specific definitions of “mammal”, but it does not take scientific training to recognize that mammals can be recognized as a group of disparate individual items that have properties that can be extended to new members of the group. This makes “mammal” a good example of a category. Categories involve individuals that are grouped together; however, critically, these are not set groups of individuals (as in, say, an individual family), but are labels that can be extended to new instances. These labels have behavioral consequences. In the context of experiments related to categorization (in the non-speech category learning literature, often artificial, lab-created categories of visual objects), these consequences might be as prosaic as which of two response buttons are pressed. For categories in the real world, the effects can sometimes be of profound import; for an animal in the wild to miscategorize a predator as prey might be a fatal mistake. Most, if not all, categories must be learned. At the very least, the precise contours of what belongs in a certain category and what does not must be fleshed out, a problem referred to as category learning. In the sections below, I discuss category learning theories inside and outside of language, with a special focus on different approaches to category learning and (especially) dual-system theories of category learning, a group of theories with growing influence outside of language and one being newly imported into the phonetic learning literature.

2.1.1 Non-Speech Category Learning

The review of modern ideas about category learning below owes much to previous summaries. The interested reader is particularly referred to reviews by Kruschke (2005, 2008), which, although slightly out of date, provide an excellent primer to some of the major schools of thought outlined below. Theories of category learning can largely be differentiated by the answers they give to two questions. First: how much abstraction is there when storing a category? Are categories described in relatively sparse terms, or is every member of each category stored? And, second, to the extent that there is abstraction, what is the nature of the abstraction?

2.1.1.1 Prototype Models

One early class of models was prototype models, which, for much of the 1970s, were ascendant in the category learning literature (Mervis & Rosch, 1981). Prototype theories, unsurprisingly, relied on the construct of a *prototype*, or ideal category member, in order to categorize items. Prototypes are generally assigned on a one-to-one basis, with each category having a single prototype. Novel items are associated with a category based on a simple, linear computation of the similarity of the new item to the prototype representing each category (Reed, 1972). For example, the category “mammal” might be defined by the most mammal-like mammal, something like a “dog”, while “bird” might be defined by the most bird-like bird, such as a “robin”. A new, potentially ambiguous animal (such as a platypus) would be assigned to the “bird” or “mammal” categories based on how similar that animal was to the prototype of each category. Platypuses would be ambiguously categorized because in some ways they resemble both dogs and robins.

Note that a prototype does not necessarily need to be a real item taken from a category; it can be an *ideal* member that may never have been witnessed by the learner. Indeed, items that are treated as prototypes are generally the quickest to be classified as belonging to their own category even when they are never directly witnessed (Homa, Cross, Cornell, Goldman, & Shwartz, 1973). They are also most likely to be retained across experimental sessions spaced days apart from each other. A delay of a week between training and testing, for example, was sufficient to lead to forgetting of individual memories of dot patterns, but not to the prototypes that those dot patterns were created from (Posner & Keele, 1970). Besides dot patterns (Homa, Sterling, & Trepel, 1981; Posner & Keele, 1968), prototype effects have also been observed for the perceived location of residence for invented biographies (Reed & Friedman, 1973) and semantic categories such as “fruit” (Rosch, 1975).

Still, prototype theory has not remained a strong force in the category learning literature. The reasons for this are fairly clear: prototype approaches to learning simply cannot accommodate the acquisition of certain types of categories. Examples can be found in Figure 2. Figure 2(a) and 2(b) show examples of categories that can easily be learned with prototypes. Larger blue and red dots in Figure 2(a) and Figure 2(b) indicate the prototypes that represent an approximate average of the items in each category, while the dotted lines in the left two figures show the spaces in each diagram that would be closer to each prototype (and would, therefore, be categorized into each category). In both cases, the prototypes can be found in the middle of the distribution of items belonging to each category. As such, any reasonable distance metric will quickly and efficiently parcel the category learning space into the relevant categories being learned;

the mischaracterization of the category boundary in Figure 2(b) is solely the result of the particular instances generated here rather than any failing of the prototype theory itself. However, Figure 2(c) shows a quite different result. In this case, the categories being learned form a bullseye, with the blue category being entirely surrounded by the red category. In this case, the prototypes in question are almost directly right on top of each other. This would lead the category space to either be divided cleanly in half, with a midpoint immediately between the two category prototypes, or to a complete failure to learn the categories, neither of which accurately matches human performance (Ashby & Gott, 1988; Ashby & Waldron, 1999). With limited exceptions (Petrov, 2011), current studies of prototype theories have generally constrained the set of situations in which prototypes might be used; some have conjectured that prototypes are only used early the learning process (J. D. Smith & Minda, 1998), giving way to other learning systems later in learning.

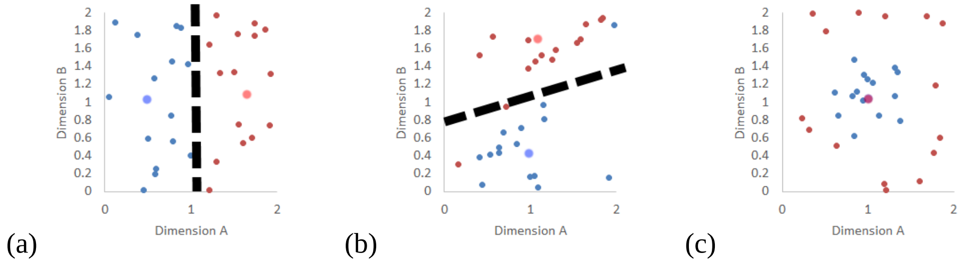


Figure 2. Hypothetical prototype category distributions

2.1.1.2 Rule-Based/Decision-Bound Models

Other early theories of category learning, building on the concept literature, often relied on explicit rules to differentiate categories. This assumes a great deal of abstraction on the part of the learner, as categories are divorced from the individual instances being taught and are instead discussed in terms of stark, black-and-white rules. These rules often resembled, say, the definitions used for ecological classifications. The category of “mammal” was traditionally described in terms of warm-bloodedness, hairiness, live birth, and milk production. If an animal checked all four of those boxes, it was assigned to the category “mammal”. But, of course, exceptions emerged; the platypus is hairy and produces milk, but also hatches from an egg. Similar exceptions in other categories are also hard to accommodate under rule-based theories, which may have hastened their demise. Strictly rule-based theories were quickly succeeded by decision-bound models, wherein categories are described by boundary conditions[1]. For “mammal”, modern approaches to ecological classification often take into account genetic similarity; one could imagine a definition of “mammal” that says a mammal is an organism that shares a certain proportion of its genetic code with already-existing members of that category. Platypuses would then be sorted into the mammal group because they are located on the mammal side of that genetic boundary.

Decision-bound models of category learning were primarily advanced in the late 1980s and early 1990s. Under decision-bound models, learners are attempting to determine the ideal boundary in perceptual space to separate multiple categories. The boundaries need not necessarily be linear, although generally under decision-bound models the boundaries proposed are subject to processing constraints that discourage overly complex boundaries (Ashby & Gott, 1988; Ashby & Townsend, 1986). Examples of distributions learnable by decision-bound theories of categorization are below in Figure 3. The distribution of Figure 3(a) is trivial to learn using a decision-bound-based model; the boundary between the categories can be given simply by a line splitting items

[1] Indeed, the similarities between these types of theories and their proponents that I largely use “rule-based” and “decision-bound” interchangeably throughout the dissertation.

in half according to their values in Dimension A. Similarly, Figure 3(b) shows a solution to categorization that requires both Dimensions A and B; again, a simple line, this time combining information from both dimensions, is used. Figure 3(c) shows the principal theoretical improvement of decision-bound theories over prototype ones: the ability to model non-linear and non-normal category boundaries. Whereas prototype theories would generally posit two entirely overlapping prototypes, some decision-bound theories (Ashby & Waldron, 1999) are happy to separate the red and blue categories using the circular boundary shown in the figure.

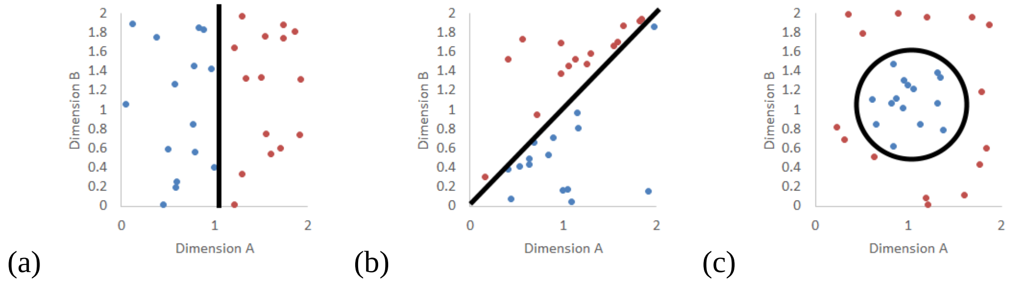


Figure 3. Hypothetical decision-bound category distributions

Decision-bound models were not without their drawbacks. Under many decision-bound models, categorization is generally deterministic. Participants classify items on one side of the boundary as solely belonging to one category, and items on the other side as belonging to another category. At times, this may accurately reflect the end state of a category learning process, but it does not adequately explain the path of acquisition before that point, as learners will, presumably, be more uncertain about how to categorize items. Even once categories are learned, it is likely that learners would still be hesitant about some aspects of categorization, particularly categorizing items close to a boundary. Decision-bound models largely lack a way to model this variability. Vandierendonck (1995) proposed resolving this by having decision bounds compete for attentional resources, with uncertainty arising from the simultaneous consideration of multiple possible decision bounds. The brittleness of decision-bound models with regard to unusual category structures, however, largely prevented their independent use; parts of these models were incorporated into later multiple-system models (see below).

2.1.1.3 Exemplar Models

The late 1980s also saw the rise of another set of category learning models: exemplar models. Exemplar models see category learning as the result of memorization of specific instances. Rather than incorporating abstraction in the form of a single category prototype or a category boundary, exemplar models eschew it entirely. Category membership is determined only by the similarity between a new item and previously observed items. The categories of "mammal" and "bird" are defined not by abstract rules or prototypes, but instead by the memories one has of animals belonging to each group: all of the specific instances of dogs, cats, horses, and humans on one hand, and all of the specific instances of robins, penguins, emus, finches, and cardinals on the other hand. Thus, a platypus might be sorted into the mammal or bird categories based on which group of animal instances best resemble the platypus.

In a way, exemplar theories represent an extreme instance of some prototype theories. While some prototype theories allowed for multiple prototypes per category, an exemplar theory of category learning resembles a prototype theory where the ratio of prototypes to items is one-to-one (Rosseel, 2002); that is, *every* item exerts a prototype-like influence on categorization. Although early reaction to exemplar theories was dismissive—they were described as "theoretically anomalous" by Mervis & Rosch (1981, p. 103)—exemplar theories came to dominate the field of category learning.

The cardinal principle of exemplar theories—that categorization is only based on observed examples of categories—emerged early. Medin and Schaffer (1978) showed that exemplar-only theories of categorization applied just as well as, if not better than, prototype theories to explaining response patterns for simple shapes and schematic faces. Probably the most widely-used exemplar theory is the Generalized Context Model (GCM) of Nosofsky (1986). According to the GCM, categorization is essentially a special class of item identification. Categorization only requires summing across how closely a new item resembles previously identified ones, using the most similar items to that new item to make a hypothesis about the category of that new item. This approach was later merged into a connectionist model known as ALCOVE (Kruschke, 1992; Nosofsky, Kruschke, & McKinley, 1992), which also incorporated information about attention on the part of the learner to different dimensions of the categories being learned. Later iterations have combined exemplars with the idea of a "random walk" (Nosofsky & Palmeri, 1997); that is, the outcome of a process of successive, iterated, random steps. Under this proposal, exemplars, when they are being used to determine the categorization of a novel item, do not have a fixed location in perceptual space. Instead, they adopt a random walk, shifting either towards the new item or away from the new item in perceptual space. When the exemplars walk into the novel item, they then contribute to the categorization of that item. Categorization occurs after that item is pushed over a pre-defined threshold of activation, once enough exemplars have made contact with the new item. This and similar approaches (Lamberts, 2000) readily accommodate reaction time patterns in category learning experiments, as both the number of readily accessible exemplars and the absence of competitors should lead the learner to respond more quickly.

What was the pull of exemplar approaches over prototype and decision-bound ones? The key is in the categories that are learnable under exemplar theories. Both prototype and decision-bound models come freighted with theoretical assumptions about the structure of categories (Ashby & Waldron, 1999). Prototype models, as outlined earlier, have significant drawbacks in learning non-linear categories. Decision-bound models are capable of learning non-linear boundaries, but often struggled in the face of uncertainty or particularly complex structures. Exemplar models have no such constraints. Indeed, barring an inability to perceptually discriminate items, exemplar models can learn most *any* category (Ashby & Alfonso-Reese, 1995; McKinley & Nosofsky, 1995), which is a strong—and testable—empirical claim, as will be examined later in this dissertation. And they can do so while accommodating some of the key theoretical insights of prototype and decision-bound theories (Medin & Schaffer, 1978). For example, MINERVA 2, an exemplar model, was shown to be capable of replicating some of the key findings related to category exemplars (Homa et al., 1973), such as the idea that category exemplars are "privileged" in processing and memory, because prototypes are often surrounded by numerous exemplars in memory. Schematic examples of category learning in exemplar models are shown in Figure 4. Each exemplar exerts a "pull" on the category space around it; areas shown in grey indicate regions where both red and blue items might influence the categorization of an item.

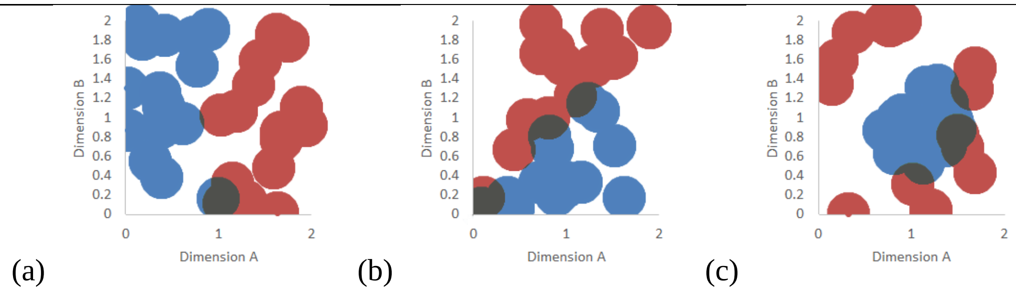


Figure 4. Hypothetical exemplar category distributions

Exemplar theories are not without criticism. Consider one of the erstwhile strengths of exemplar models, the idea that all categories are learnable, given enough training (Ashby & Alfonso-Reese, 1995). Is this actually a tractable hypothesis? Listeners are constrained in their category learning, failing to learn some complex category structures even after many days of training (McKinley & Nosofsky, 1995) and preferring simple, linear category boundaries to more complex ones (Ashby, Waldron, Lee, & Berkman, 2001). Exemplar models also find the learning of hierarchies to be quite challenging; for example, monotremes and primates are examples of mammals, and mammals are examples of animals. Yet the fact that different stimulus dimensions play a role in determining category membership at each level of a hierarchy is challenging to accommodate under a theory that only depends on item memorization, as every level of the hierarchy must be memorized simultaneously, with different attributes of the stimuli weighted differently for each level (Lassaline & Murphy, 1998). Finally, exemplar models of category learning rely strongly (perhaps even more strongly than other theories of category learning) on notions of similarity; yet how "similarity" is assessed in a model that only has access to individual memories, and how a learner selects which attributes of a stimulus can be used to calculate it, are both highly underspecified (Kruschke, 2005).

2.1.1.4 Multiple-System Models

With all three classes of model previously outlined showing some weaknesses, one might wonder whether any one system can explain category learning to a sufficient extent. For one class of researchers, the answer is, clearly, “no”. Multiple-systems theories emerged in the late 1990s as a significant competitor to exemplar theories, perhaps bolstered by claims about the mathematical interchangeability of exemplar, decision-bound, and prototype theories (Ashby & Maddox, 1993; Rosseel, 2002). An example of this interchangeable nature was sketched at the beginning of the section on exemplar theories; an exemplar theory is equivalent to a prototype theory with a single prototype for every instance of a category. As of this writing, it does not seem to be much of a stretch to describe multiple-system theories as being the most popular in the contemporary field. Even some of the strongest proponents of exemplar models of categorization have begun to acknowledge the possibility of multiple systems of category learning, arguing that making use of many possible category learning models shows that there is “clear evidence that participants learn to categorize in more than one particular way” (Donkin, Newell, Kalish, Dunn, & Nosofsky, 2015, p. 945).

The reasons for this shift are many, but two stand out. First, inter-learner variability. Many of the foundational models in the exemplar model literature were based on data aggregated across participants. When investigating participants on an individual-by-individual basis, some participants show much more rule-like behavior than would be predicted on the basis of exemplars alone (Kalish & Kruschke, 1997). Changes between the use of exemplar-based learning and rule-based learning are strongly associated with different task demands and experimental contexts, which would be surprising if only a single system determined learning (E. E. Smith, Patalano, & Jonides, 1998). A second relates to category generalization. Exemplar theories predict that exceptions to simple rules might have a strong effect on categorization, particularly if they are present in an environment relatively clear of other items. Instead, participants almost entirely ignore those isolated exceptions, categorizing the vast majority of nearby items as belonging to the rule-described category (Erickson & Kruschke, 2002; E. E. Smith et al., 1998).

If multiple systems are at work in category learning, then, what are their forms? Theoretically, any combination of the three systems described above could fit category learning behavior better than a single one of them. However, to my knowledge, no theories have attempted to fuse all three approaches into a single theory, or to combine prototypes and decision bounds. Although some models have combined exemplars and prototypes (Homa et al., 1981; Storms, De Boeck, & Ruts, 2001), detailed analysis of such approaches has in general found that the exemplar-prototype hybrids do not fit category learning data any better than models that include exemplars alone (Busemeyer, Dewey, & Medin, 1984).

Most multiple system approaches, then, take a tack that combines an exemplar-like system with a rule-like system. Under RULEX (RULes and EXceptions; Nosofsky & Palmeri, 1998; Nosofsky, Palmeri, & McKinley, 1994), learners first attempt to sort items into categories according to simple, linear rules. If this is not successful up to a certain prespecified proportion of trials, the rule-based system falls back first on a lower threshold of “success”, and finally on conjunctive rules (rules that require an “and” to spell out). Meanwhile, once the accuracy of categorization reaches a certain level, the learner then seeks out exceptions to the rules that were learned, using an exemplar-like process to assign items to those exceptions. ACT-R (Anderson & Betz, 2001) combines the rule-learning system of RULEX with the exemplar-based system of the exemplar-based random walk model (Nosofsky & Palmeri, 1997) within a connectionist framework that has been applied to other cognitive tasks. Under ATRIUM (Attention To Rules and Instances in a Unified Model; Erickson & Kruschke, 1998, 2002), novel items are evaluated by both rule-based and exemplar-based modules to determine categorization; the reliance on each module is contingent on attentional modulation.

The most successful dual-system model, COVIS (COmpetition between Verbal and Implicit Systems; Ashby, Alfonso-Reese, Turken, & Waldron, 1998), combines a familiar rule-based system with a second decision-bound system, albeit one that largely replicates exemplar-based learning. Like in ATRIUM, the rule-based and similarity-based systems compete to determine the response that is output by the category learning system. The rule-based system makes explicit, verbalizable hypotheses about the categories being learned over the course of the experiment. The similarity-based system, meanwhile, is not explicitly an exemplar one. It is instead based on complex decision bounds that are not restricted in their shape or number (Ashby & Waldron, 1999) inspired by the firing of dopaminergic basal ganglia neurons. According to COVIS, at the onset of learning, learners tend to rely more on their rule-based systems; over time, they fall back on their similarity-based system to assemble detailed category structures. Platypuses would first be assigned to “bird” or “mammal” based on abstract properties, but after further exposure a similar categorization decision would instead be resolved on the basis of similarity.

COVIS and similar multiple-system accounts of category learning are intriguing in part because of their interest in and reliance on neurobiological evidence, as well as their specific behavioral predictions. For example, tasks that are said to require the rule-based system are taxed by limits on the memory capacities required to make hypotheses about category membership (Waldron & Ashby, 2001). Meanwhile, tasks that require the similarity-based system are easily thrown off by changes to feedback that throw off the dopaminergic reward circuitry that is said to underlie the memorization of the decision bounds that make up the similarity-based system (Ashby & Maddox, 2005; Maddox & Ashby, 2004). Having two category learning systems that are set up in this way can lead to some unexpected predictions that have later accrued significant evidence. For example, patients with Parkinson's disease tend to show deficits in tasks related to similarity-based learning attributed to deficits in reward circuitry (Shohamy, Myers, Kalanithi, & Gluck, 2008). Furthermore, non-human animals (in one particular case, rats) can outperform human on tasks that require exemplar-based learning due to their weak abilities with regard to rule-based learning when compared to humans (Vermaercke, Cop, Willems, D'Hooge, & Op de Beeck, 2014).

This neurobiological specificity is likely one of the main reasons for the embrace of multiple-system theories. Although exemplar-only approaches have been linked to some speculation about, say, the role of the hippocampus in category learning (Pickering, 1997), these connections are much more tenuous than those proposed by multiple system theorists. Indeed, multiple category learning systems appear to have been embraced wholesale by the cognitive neuroscience community, particularly bolstered by perceived connections to hypotheses about multiple memory systems (Poldrack & Foerde, 2008;

Seger & Miller, 2010; E. E. Smith & Grossman, 2008). Some have gone as far as to say that "it is simply not possible to maintain a single-system approach to learning and memory if one takes neurobiology seriously" (Poldrack & Foerde, 2008, p. 203).

Multiple system approaches have not been without their detractors. Indeed, many studies used to promote COVIS, such as the one that suggested that dual task paradigms seem to tax categories that are best learned using the rule-based category learning system (Waldron & Ashby, 2001), have been followed by exemplar-only rebuttals (Nosofsky & Kruschke, 2002), and even multiple-system ripostes to the rebuttals (Ashby & Ell, 2002). Such back-and-forths have largely centered on small details of the decisions made during the modeling or the paradigm used to collect the data.

Borrowing inspiration from broad-based criticisms of dual-system models in psychology (Keren & Schul, 2009), however, wider criticisms have been leveled against dual-system models (Newell, 2012; Newell, Dunn, & Kalish, 2011). These broad criticisms have focused on two parts of the dual-system enterprise: first, the utility of the behavioral and neuropsychological dissociations used to support the idea of two systems, and, second, the relevance of neural findings to the proposed division of category learning into systems. On the first front, Newell et al. (2011) convey skepticism about the utility of dissociations for determining the presence of non-shared neural or behavioral resources, even the vaunted "double dissociation" (neural or otherwise). Consider the dual-task paradigm briefly alluded to above. In that paradigm, the simplicity of the category structure to be learned (simple vs. complex) is crossed with the task demands (single-task vs. dual-task). The dual-task paradigm is much harder than the single-task one for the simple category structure, but the two tasks are not as strongly differentiated in the complex task structure. That is in line with a multiple-system paradigm in which the different category structures would be processed by different category learning systems. But it is also in line with a single-system account in which performance in the complex condition is subject to some sort of floor effect; the task demand would simply not be able to make a difference when performance was already at a very low level.

The second critique, that of the relevance of neural findings, relates to Marr's (1982) descriptive hierarchy. According to Marr, problems in cognitive science can be described at three levels. A *computational* description of a problem lays out what processes need to be executed and why those processes are important. An *algorithmic* description details the representations that are necessary in order to carry out the process in question, as well as the transformations that are required in order to turn inputs into outputs. And an *implementational* description describes the physical mechanism that carries out the algorithm. In cognitive science, this is almost certainly something in the brain. In fact, COVIS (Ashby et al., 1998) was originally formulated with each of these levels described. However, Newell (Newell, 2012; Newell et al., 2011) argues that some of the arguments brought forth to support multiple-systems approaches inappropriately conflate the levels of description in Marr’s hierarchy, in particular by using neural metrics to inform the computational theory being advanced.

Many of the same arguments regarding dissociations also apply to neural measures. Associations between certain tasks and certain brain regions might easily reflect a single-system account of learning (as mediated by, say, difficulty). Even over and above that, though, it is useful to consider to what extent neural measures *should* inform computational theories of a problem. The idea that multiple, distributed brain regions underlie category learning does not necessarily mean that multiple cognitive systems are at work. Evidence for a split at the implementational level should not imply evidence for a split at the computational level. These bigger-picture issues remain challenging for proponents of multiple systems.

2.1.1.5 Rational Models

A final class of models, often termed rational models, provides a more computationally-focused account of category learning. Rational models attempt to summarize the ideal end-state of learning rather than the actual procedures that are implemented, as generally discussed by the models above (Tenenbaum, Griffiths, & Kemp, 2006). For rational models of category learning, categorization is derived by combining information about the likelihood of seeing an object with a particular set of features given the category being probed with the prior probability of choosing a particular label (Sanborn, Griffiths, & Navarro, 2010). For example, the mislabeling of platypuses may have resulted from the very low likelihood, given other mammals seen in the world, of seeing an egg-laying mammal. Categorization is also affected by the prior probability. A similar fuzzy, egg-laying animal being discovered in, say, New Zealand would also be unlikely to be labeled as a mammal (at least at first) because New Zealand does not have any native, ground-based mammals, thus making a zoologist's prior likelihood of using the "mammal" label to refer to anything to be quite low.

Using a rational approach to category learning has some interesting consequences. For example, both prototype and exemplar models can be tied in with rational approaches, with the differences between the models being described in terms of the ways that the relevant probabilities are computed (Ashby & Alfonso-Reese, 1995). That said, however, it is not common for rational models to be tied to the algorithmic level that many of the previously-outlined models operate at. One early attempt to do this came from Anderson (1991), who implemented a system of cluster assignment intermediate between prototype and exemplar models as a way to model the probability of observed features given a single category. However, the model was extremely complex, requiring the evaluation of hypotheses corresponding to the assignment of *every* exemplar to *every* possible combination of clusters; it also was strongly order-sensitive. Later rational approaches attempted to rectify these issues by suggesting different ways to assign and update the predicted probabilities (Sanborn et al., 2010). Although this dissertation is focused primarily on the algorithmic level, disambiguating between the predictions of the theories outlined in the previous sections, rational approaches can illuminate interesting corners of the present approach. For example, speculation about "over-hypotheses"—hypotheses about hypotheses, or constraints on the types of ideas that can be entertained about categories—can benefit from a hierarchical Bayesian approach to learning (Kemp, Perfors, & Tenenbaum, 2007).

2.1.2 Speech Sound Categorization

Categorization has been a problem often discussed in the realm of speech perception, given what has been termed the "invariance problem" in speech. Speech is, by its nature, variable. Speakers differ in age, gender, native speaker status, and dialect, all of which have consequences for the sounds they produce. Even within a single speaker, speech sounds differ according to the adjoining speech sounds they are produced with (through coarticulation), to the prosodic context in which they are found, to the emotional and physical state of the speaker, and to chance in line with noise in the speech production system. Yet, despite all of those sources of variability, listeners are still capable of understanding the speech of a wide variety of speakers across a wide variety of contexts. The question arises, then: how do listeners cope with the variability present in the signal, such that they place the tokens into the correct category? Previous proposals along these lines have differed in terms of what attributes the perceptual system must incorporate in order to promote invariance.

2.1.2.1 Rule-based Models

Traditional linguistic approaches emphasize the merits of using "distinctive features" to differentiate phonetic segments. Listeners are said to employ a set of (perhaps innate) features that align with certain important contrasts in the languages they use. These features—or, at least, their auditory correlates—then allow listeners to abstract away from the tokens of individual sounds and understand the signal. The ascription of features to sounds is generally said to follow along the lines of a boundary or rule; for example, a stop in English might be classified as "voiced" if it has a voice onset time (VOT) of 35ms or smaller, and "voiceless" if it falls above that boundary. The idea of phonetic categories being the result of an orderly combination of distinctive features is at its peak in the International Phonetic Alphabet (Ladefoged & Halle, 1988), where, say, the [p] sound is described in terms of its values on three categorical dimensions: voicing (voiceless), place (alveolar), and manner (stop).

Early evidence for the value of distinctive features came from studies of short-term memory, in which participants who learned lists of monosyllables that shared certain distinctive features showed more confusion at recall than participants who learned lists of monosyllables that did not share distinctive features, for both vowels (Wickelgren, 1965) and consonants (Wickelgren, 1966). Researchers have looked for consistent acoustic correlates of these distinctive features, seeking invariant cues for, for example, place of articulation (Stevens & Blumstein, 1978). Invariance has been ascribed to various aspects of the speech signal. One early explanation suggested "template-matching" of the signal to particular distinctive features (Blumstein & Stevens, 1979). Other proposals have included constant ratios of certain acoustic parameters that distinguish consonant places of articulation under the umbrella of "locus equations" (Sussman, McCaffrey, & Matthews, 1991) and measures that discriminate between places of articulation according to the moments (i.e., measures of distribution, such as mean, skew, and kurtosis) of the distributions of consonants produced at each place (Forrest, Weismer, Milenkovic, & Dougall, 1988).

Boundary-based theories of category learning have continued to see use in the phonetics literature. The finding that Korean-speaking people hearing speech sounds along an alveolar stop continuum show no evidence for a specific, category-related magnetoencephalography (MEG) component when Russian-speaking people do, for example, has been credited to the presence of a category boundary for Russian speakers that is absent for Korean speakers (Kazanina, Phillips, & Idsardi, 2006). Distinctive features have been used to create an abstract lexical representation of words that can interface with higher-level (e.g., syntactic) linguistic domains (Poeppel et al., 2008). Yet many of the theories that use distinctive features assume that representations are boundary-based rather than directly testing that idea.

2.1.2.2 Prototype Models

As in the general category learning domain, it has long been recognized that some instances of phonetic categories are “better” than others (J. L. Miller, Connine, Schermer, & Kluender, 1983). Although items with large VOTs are generally characterized as voiceless, a stop with a 300ms VOT will be perceived to be a “worse” [p] more often than a stop with a 60ms VOT. Listeners are perfectly happy to ascribe goodness ratings to phonetic categories. As will be discussed in much greater detail in Section 2.3, these ratings are dependent on the rate of speech in a way that parallels shifts in category boundaries based on speech rate (J. L. Miller, 1994, 1997; Volaitis & Miller, 1992). They are also subject to other, parallel relationships with different acoustic cues to the presence of certain phonetic features. The presence of [t] in the word *stay*, for example, is cued by both the duration of silence before the onset of the following vowel sound and first formant transition of that vowel. Shifting the frequency of the F1 transition higher leads listeners to require a longer silence to perceive a single token as a “good” example of *stay* (Hodgson & Miller, 1996).

The existence of perceptual distinctions between items within categories has led some to posit prototype theories of phonetic category learning. Samuel (1982) found that repeatedly playing a prototypical [g] sound led to participants to later classify sounds ambiguous between [g] and [k] as a [k], indicating that prototypical [g] sounds were more likely than non-prototypical [g] sounds to influence participants’ later perception of ambiguity. Borrowing explicitly from the prototype category learning literature (Mervis & Rosch, 1981), Samuel (1982) suggested that this implied categories were represented as single prototypes. Later studies suggested that infants also made use of prototypical representations, with category prototypes (as assessed by adults) leading to significantly broader generalization on the part of 6-month-olds (Grieser & Kuhl, 1989). The latter effects were connected to a “perceptual magnet effect”, the idea that category prototypes have a special role in the organization of categories that is species-specific; for instance, the sorts of generalization findings present for adults and infants do not extend to monkeys (Kuhl, 1991). Later modeling indicated that the reason for these findings may be in part because items near the category prototype were less discriminable than those further away (Iverson & Kuhl, 1995). These findings collectively were taken to suggest that prototypical category members have a privileged status in speech perception, and that that privileged status is indicative of the way that phonetic categories are stored.

However, these findings have been questioned. Some of this has been methodological. The setup of the perceptual magnet effect assumes that the non-prototypical vowels used were still perceived as belonging to the vowel categories being tested, an assumption that was questioned in follow-up studies (Lively & Pisoni, 1997). More importantly, however, the idea that some phonetic category tokens can be perceived as better examples of that category than others (or even as “the best”) can be explained by models other than prototype categories of category learning. Under exemplar theories of category learning, an item might be perceived as the prototypical instance of a category because it is surrounded by many exemplars belonging to that category (Lacerda, 1995). A good [g] is a good [g] because it is surrounded by many [g] tokens, not because it is somehow “special”. Under rational models of phonetic category learning, the perceptual magnet effect is explained by a listener’s certainty about a speaker’s *intended* production (Feldman, Griffiths, & Morgan, 2009). As discussed previously, even decision-bound models could explain some of the patterns of category goodness ratings, given a degree of uncertainty about the location and type of boundary present in the signal (Vandierendonck, 1995). The prototype models of categorization thus depend on what is essentially a non-sequitur; just because there are prototypical category members does not not mean that the categories are represented using prototypes.

2.1.2.3 Exemplar Models

Phonetic category learning has also been the home of exemplar-based theories. One particularly well-cited instance of an exemplar-based theory of phonetic perception is that of Pierrehumbert (2003). Under Pierrehumbert's (2003) model, speech sound categories are simply the collection of multiple memorized pairings of individual speech sound tokens (i.e., exemplars) and categories. New items that are fed into the system are simply compared to previously observed ones. The categories that the most similar previous items belong to are compared with one another, and the new item is paired with the category that has the most (and most similar) category connections. Under exemplar-based theories, "memory capacity is large…representations in memory are extremely detailed, and…[exemplars] include time and many other nonspeech properties" (Pierrehumbert, 2016, p. 10.4). The [p] category, then, is defined by the many specific instances of the [p] sound that have been encountered on the part of a listener; distinctive features do not play a role in categorization. These theories have been explicitly inspired by exemplar-only theories of category learning, especially the GCM (Nosofsky, 1986) and MINERVA (Hintzman, 1986; Homa et al., 1973).

Much of this speculation arose from discussion of "speaker normalization", the idea that listeners must find a way to rid the signal of, say, speaker-specific acoustic properties that are not directly linguistically-relevant in order to process a signal. Such speaker normalization is generally assumed by many theories of lexical access (McClelland & Elman, 1986), and can be seen as the output of rule-based and prototype-based theories of phonetic categorization. However, the idea of speaker normalization is not above criticism. Items from word lists are better and more quickly recognized when both trained and tested using a single speaker, for example (Palmeri, Goldinger, & Pisoni, 1993). Such findings are expected under exemplar theories, in which *all* acoustic information, including that necessary to distinguish between speakers, is saved in memory. Under these theories, even very small phonetic details describing the differences between sounds can be critical, as one's recollection of these fine phonetic details may take critical importance in distinguishing between categories (Hawkins, 2003). Exemplar theories draw on a rich tradition of variation between speakers, dialects, and languages within the phonetics literature. They have been used to explain why diachronic and synchronic sound change is often dependent on the lexical frequency of a word (Bybee, 2002) and why listeners are sensitive to within-category phonetic variation in lexical access (McMurray, Tanenhaus, Aslin, & Spivey, 2003).

This represents a radical departure from the previously-described classes of theories, which all involve the assembly of abstract categories through defining properties. The authors of such theories have not been shy about the strong departure of their hypotheses from previous orthodoxy. Hawkins (2010), for example, analogized the perception of abstract categories in language to other "auditory illusions". Other exemplar-based theorists have argued "against formal phonology" (Port & Leary, 2005) or in favor of going "beyond phones and phonemes" (Port, 2007) due to the "quixotic" nature of quests for individual units (Goldinger & Azuma, 2003). These theories have also been applied towards ideas of lexical access, with exemplar-based theories arguing against abstract lexical forms entirely (Goldinger, 1998). Such rhetoric, though, obscures the notion that a recognition of the proper importance of variability and speaker-dependence does not necessarily require jettisoning the idea of abstract categories entirely. One way to embrace variation without removing abstraction is to adopt a multiple-system model of phonetic learning.

2.1.3 Multiple-system Models in Language

Multiple-system models are also present in language. Indeed, they have been proposed at a variety of levels of analysis; the acquisition of morphosyntax, lexical items, and phonetic categories have all been approached using multiple-system models. However, such efforts have usually been conducted in parallel, without cross-talk between the divisions nor frequent reference to dual-system models elsewhere. Below, I sketch out some of the principal multiple-system theories of language learning, with reference to commonalities between them as well as shared properties with cognitive psychology theories of category learning described in Section 2.1.1.4. It is important to note that these models do not always work along the same lines as ones of category learning. It is unlikely that the “categories” being learned in morphosyntax are perceptual in nature, nor do the categories used in word learning always neatly reflect items that can be perceived. Still, these models provide an informative model for how dual-system views of language might work. And, in many ways, they show some intriguing similarities to newly-proposed dual-system models of phonetic category learning that will be subsequently discussed.

2.1.3.1 In morphosyntax

Morphosyntax provides the most fully-realized example of a dual-system model within language. Its origins can largely be traced to an influential critique (Pinker & Prince, 1988) of connectionist approaches to the acquisition of the English past tense (specifically, Rumelhart & McClelland, 1986). Connectionist approaches to syntax bear many similarities to exemplar-only ones, particularly in their avoidance of abstract representations. In their critique, Pinker and Prince (1988) pointed out numerous potential flaws in the connectionist implementation of past tense formation in English, especially related to instances in which the model ignores forms that traditionally are said to require knowledge of lexical representations (e.g., *ring* and *wring* being homophones with different past tenses, or the past tense of *grandstand* being *grandstanded*, not *grandstood*). This critique was later parlayed into a dual-system model that explicitly distinguishes between rules that involve a productive generation process and words that require memorization. Irregular verbs are past tense items that are memorized just like any other lexical item (Pinker, 1998; Pinker & Ullman, 2002). This idea was later parlayed into a popular science book named, appropriately enough, *Words and Rules* (Pinker, 1999). Although the main topic of discussion here was the English past tense, it was argued that this distinction could also apply to other languages and processes, such as pluralization in Hebrew (Berent, Pinker, & Shimron, 2002).

Just as with category learning, evidence for dissociations between “word” and “rule” systems mounted (Lavric, Pizzagalli, Forstmeier, & Rippon, 2001). Behaviorally, for example, irregular past tense forms show a great deal of variation in their acceptability ratings, while regular past tense forms show little systematic variation, suggesting that the two classes of past tense verbs are underlain by two separate systems (Ullman, 1999). Neuropsychologically, patients with declarative memory deficits (such as Alzheimer's) produced more errors when producing irregular past tense forms than when producing regular ones, while patients with procedural memory deficits (such as Parkinson's) showed the opposite pattern of errors (Ullman et al., 1997). In an event-related potential (ERP) design, ungrammatical, regular past tense verb forms yielded a left-lateralized anterior negativity (LAN) that did not show up for errors of irregular morphology (A. J. Newman, Ullman, Pancheva, Waligura, & Neville, 2007).

Ullman and colleagues have used these dissociations to argue for a dual-system account of morphosyntactic learning (Ullman, 2004, 2016). Under this account, they propose a distinction between aspects of grammatical competence that need to be explicitly memorized and aspects that require a rule. These two systems are linked to declarative memory and procedural memory, respectively. Declarative memory is associated with conscious recollection and explicit learning; procedural memory is linked to routinized sequence learning and implicit learning (Squire, 2009). Rule-like grammatical acquisition in a first language, for instance, has been correlated with scores of procedural-memory-linked implicit learning, not declarative-memory-linked explicit learning (E. Kidd, 2012). Other connections come from neuroimaging evidence. During decisions about grammaticality over the course of artificial language learning, procedural learning regions, such as the basal ganglia, tend to be active when compared to other, non-grammatical decisions (Petersson, Folia, & Hagoort, 2012). Explicit learning and implicit learning in an artificial language paradigm have been associated with different patterns of resting-state functional connectivity (Yang & Li, 2012).

Dual-system models of syntactic learning have additionally been applied to populations that often show language deficits, including second-language learners and those with language disorders (especially Specific Language Impairment, SLI). According to Ullman (2001), second-language acquisition is typically guided by declarative memory, in line with the typical bent of the adults learning the second language. However, *successful* L2 acquisition depends on a switch of control over the learning process to the procedural memory system, which can start to instantiate rules, a switch that mirrors that proposed in non-linguistic category learning (Ashby et al., 1998). Second-language Spanish speakers, for example, might use their declarative system to memorize *every* verb conjugation, regardless of its regularity, while Spanish speakers might only memorize some forms (Bowden, Gelfand, Sanz, & Ullman, 2010). The switch necessary to turn non-native speakers into native-like ones might be aided by implicit, immersion-like training (Morgan-Short, Sanz, Ullman, & Steinhauer, 2010; Morgan-Short, Steinhauer, Sanz, & Ullman, 2012).

Language disorder can also provide a fruitful avenue of study, not in the least because it appears that declarative memory can be used as a “work-around” for people with a disorder who are processing language (Ullman & Pullman, 2015). Although the use of the dual-system approach to syntactic processing has been tested briefly in the dyslexia literature (Hedenius et al., 2013), the primary application has been to studies of SLI, which Ullman and colleagues has argued stems primarily from a procedural memory deficit (Conti-Ramsden, Ullman, & Lum, 2015; Lum, Conti-Ramsden, Page, & Ullman, 2012).

2.1.3.2 In word learning

Although dual-system theories of morphosyntax are generally predicated on a split between “words” and “rules”, dual-system accounts have also been proposed solely within the domain of “words”, centering on word learning in particular. Although, these theories are not explicitly tied to categorization, they provide an illustrative example of a two-system model of learning in language. Unlike the procedural/declarative memory distinction used in morphosyntax, dual-system accounts of word learning (Davis & Gaskell, 2009; Lindsay & Gaskell, 2010) instead borrow from “complementary learning systems” accounts of recognition memory (Norman & O’Reilly, 2003) that divide recognition into two components, a hippocampal component and a neocortical component (though see Ripollés et al., 2014 for evidence that word learning may also depend on the basal ganglia). Under dual-system accounts of word learning, learning is separated into two components: a hippocampal one, which rapidly familiarizes learners with new words, and a neocortical one, which helps stabilize and preserve the recently-learned words, their meanings, and their relationships with similar words. This is in some ways similar to distinction between “lexical configuration” and “lexical engagement” (Leach & Samuel, 2007), with “lexical configuration” referring to knowledge of, say, the meaning of a word and its spelling, and “lexical engagement” referring to real-time integration of the word in later processing (as in, say, lexical competition, or perceptual learning).

Much of the behavioral evidence for dual-system models of word learning comes from studies of sleep consolidation. It is well known that sleep promotes memory consolidation (Stickgold, 2005). According to complementary learning systems accounts, consolidation is necessary to promote the activity of the neocortical system in stabilizing a new word. As such, dual-system theories of word learning predict that, while information about the form and meaning of a word might be gained quickly using the hippocampal system, its ability to compete with items in its lexical neighborhood (i.e., items that are similar to it phonetically) requires sleep. Indeed, that is generally what is found across a variety of paradigms that have made use of lexical competition (Dumay & Gaskell, 2007; Gaskell & Dumay, 2003; but see Lindsay & Gaskell, 2013). Novel words become more "word-like" over time. For instance, real words lead to greater inhibition of their lexical neighbors when they share onset syllables; novel words only do so after sleep consolidation (Dumay & Gaskell, 2012). This also translates over to eyetracking paradigms that assess real-time lexical competition (Magnuson, Tanenhaus, Aslin, & Dahan, 2003; Wang et al., 2016) and to affixes as well as words (Tamminen, Davis, Merkx, & Rastle, 2012). 7- to 12-year-old children also benefit from sleep consolidation, showing similar effects of competition as adults (Henderson, Weighall, Brown, & Gaskell, 2012). On the other end of the spectrum, the age-related decrease in the duration and quality of sleep has been linked to deficits in word learning in older adults (Kurdziel, Mantua, & Spencer, 2016).

The behavioral dissociations discussed above have been supplemented with neuropsychological study. Lesions in the left medial temporal gyrus, an area associated with the neocortical word learning stream, have been strongly linked to deficits in the comprehension of single words (Dronkers, Wilkins, Van Valin, Redfern, & Jaeger, 2004). Meanwhile, damage specific to the hippocampus, as might occur after certain brain injuries during childhood (Vargha-Khadem et al., 1997) has also been linked to word learning deficits, although continued neocortical connections can sometimes lead people with hippocampal damage to maintain some level of intact word learning (Gadian et al., 2000; Gardiner, Brandt, Baddeley, Vargha-Khadem, & Mishkin, 2008), perhaps due to continued neocortical connections. Dual-system theories also show promise for illuminating other language disorders, and especially the treatment of those disorders, although such research is still in its infancy (Storkel, 2015).

Adults without brain injuries also show patterns that suggest that two systems are at play. One study using EEG found a distinction between explicit representations of word meaning, which developed quickly, and implicit representations, which took much more time (Batterink & Neville, 2011); the "explicit representations" could be connected to the hippocampal system of learning, while the "implicit representations" could be tied to the neocortical one. In fMRI designs, hippocampal activation spikes at the beginning of a word learning paradigm and later trails off (Paulesu et al., 2009). Meanwhile, activation in the neocortex to novel words does not become similar to that of real words for quite some time, though sleep consolidation seems to help the process along (Davis, Di Betta, Macdonald, & Gaskell, 2009; Orfanidou, Marslen-Wilson, & Davis, 2006).

This approach is not without its skeptics. Most of those skeptics insist that word learning is indeed extremely fast and automatic, citing both neural (Shtyrov, 2012; Shtyrov, Nikulin, & Pulvermuller, 2010) and behavioral (Kapnoula & McMurray, 2016; Kapnoula, Packard, Gupta, & McMurray, 2015) data. Most notably for the present work, the dual-system account of word learning is not generally focused on the *algorithms* that go into word learning in a way that the syntactic or category learning literatures are.

Representations, whether abstract or word-specific, are not often discussed. This sets dual-system accounts of word learning apart from the rest.

2.1.3.3 In phonetics

Given the importance of categorization for both category learning and for language, it is unsurprising that dual-system models of category learning have begun to be applied to the perception of phonetic categories (Chandrasekaran, Koslov, & Maddox, 2014). The primary approach has been to essentially borrow the dual-system approach in visual category learning to speech perception wholesale, complete with a frontal, rule-based system and a basal-ganglia-dominated, similarity-based system trading off against each other. As before, category learning tends to be dominated by the rule-based system early in the learning process, before control is largely passed to the similarity-based system. Although this need not necessarily be the case—one dual-system model of phonotactic learning (Moreton, Pater, & Pertsova, 2017) involves a maximum entropy framework—it is the case that the dual-system theories of Chandrasekaran and Maddox are, at the moment, the most clearly elucidated.

Many of the experiments assessing dual-system models of phonetic learning have used lexical tone learning from languages such as Mandarin Chinese to examine how listeners might use rule-based and similarity-based category learning systems to pick up speech sounds. Chandrasekaran, Yi, and Maddox (2014) used native English speakers to examine the acquisition of Mandarin tone categories. They found that listeners were faster to learn Mandarin tone categories under conditions that enhanced the exemplar-based category learning system, including providing immediate and simple feedback during the experiment. Computational modeling of the acquisition of Chinese tone categories echoes the idea that similarity-based learning is particularly important to speech sound categories, with learners switching over to a similarity-based approach to learning Mandarin tones generally learning better than those who avoided doing so (Maddox & Chandrasekaran, 2014).

Just as in the cognitive psychology literature, this has spurred interest in the language domain about the role of dopaminergic reward circuitry on speech sound learning (Lim, Fiez, & Holt, 2014). This speculation has particularly focused on the basal ganglia, a subcortical structure that contains much of the dopamine-related circuitry in the brain, including regions that are impaired in classic dopamine-related disorders (e.g., Parkinson's and Huntington's disease). The basal ganglia are said to be of crucial importance to the similarity-based category learning system, since it is dopamine that underpins the assimilation of feedback to the category learning situation. People with basal ganglia-related disorders therefore are said to be impaired on their acquisition of categories that require the use of similarity-based learning (Shohamy et al., 2008). fMRI studies of speech sound category learning echoed findings from the non-linguistic category learning literature with similarity-based learning associated with greater activation in the basal ganglia (Yi, Maddox, Mumford, & Chandrasekaran, 2016).

Phonetic category learning has also been studied with a variety of other interesting populations. Older adults, for example, have been shown to have deficits specifically related to the rule-based system of category learning, associated with deficits in working memory more generally (Maddox, Chandrasekaran, Smayda, & Yi, 2013). Musicians are generally faster at switching to a multidimensional learning strategy when learning new categories of speech sounds, perhaps indicating a greater flexibility in their rule-based category learning systems than non-musicians (Smayda, Chandrasekaran, & Maddox, 2015). Meanwhile, holders of a single nucleotide polymorphism (SNP) variant of the FOXP2 gene—a gene intimately associated with normal language functioning, with unrelated polymorphisms in this gene leading to massive and specific deficits in language functioning (Enard et al., 2002)—switched more quickly to a similarity-based learning strategy in the type of lexical tone learning task that was sketched briefly earlier, in which such a strategy is optimal (Chandrasekaran, Yi, Blanco, McGeary, & Maddox, 2015), a finding perhaps related to earlier findings that the same polymorphism also seems to modulate frontal activation during a simple reading task (Pinel et al., 2012). A dual-system approach can also be used to inform knowledge of what are typically said to be language-related disorders: people with dyslexia exhibit difficulty in non-linguistic auditory category learning in situations that are said to tax exemplar-based learning systems (Gabay & Holt, 2015).

Some of the most striking pieces of evidence in support of dual-system models relate to the idea that phonetic category learning is most strongly associated with the similarity-based learning system. This leads to some strongly counterintuitive predictions about individual differences in learning. Children with better working memory scores found it more difficult than children with worse working memory scores to learn talker categories, perhaps because they are erroneously relying on a rule-based strategy for a task in which a similarity-based one is more appropriate (Levi, 2015). Conversely, people with elevated depressive symptoms are *better* at speech learning tasks than people with lower depressive symptoms, as their generally-elevated cognitive load makes it less likely that they will use a rule-based strategy to learn new speech sounds (Maddox et al., 2014).

2.1.4 Summary

As should be clear, debates about the nature of category learning inside and outside of language often parallel each other. For example, prototype theories had a brief period of popularity in both literatures before being eclipsed by other conceptions of categories. Exemplar theories have received sustained attention for at least a couple of decades. Yet there are some points of divergence. Rule-based models continue to see use in the speech perception literature, despite their replacement with dual-system theories outside of language. Yet in this, too, the speech perception literature appears to be following the trend of the cognitive psychology literature, echoing theories proposed in morphosyntax and in word learning.

What do these findings imply for the debate about modularity? Certainly, if the exact same systems underlie the acquisition of categories in all circumstances, category learning is not an area in which phonetic perception is modular. This seems to be the solution favored by almost all proponents of exemplar-only phonetic learning (Hawkins, 2010; Pierrehumbert, 2016; Port, 2007), although it should be noted that this does not *need* to be the case even under those views. For example, learners may start with "pre-programmed" speech-specific exemplars that influence their categorization of items at a young age, or the dimensions used for speech category learning could be speech-specific. Still, speech specificity may be more easily accommodated by a framework in which listeners form abstract categories, as in a rule-based system or a multiple-system models with rules. Learners could, say, have speech-specific restrictions on the types of hypotheses that could be entertained within the rule-based system. Not many studies, though, have directly compared the acquisition of speech sound categories with the acquisition of other categories.

2.2 Speech Category Learning

Based on the previous review, it is clear that phonetics is an interesting test case for theories of category learning. Phonetic categories are multifaceted, complex, and extremely important for effective language functioning. Previous attempts to bring non-linguistic category-learning theories into the field have led to a richer understanding of language. Yet there are also reasons to think that phonetic category learning might work differently from non-linguistic category learning. Most theories of category learning were formulated in the visual domain, where learning may proceed along different lines. Both motor theories of speech perception (Liberman, 1982) and domain-specific auditory theories of speech perception (Remez, 1989) would suggest that phonetic objects are evaluated differently from other things in the world, including visual objects. Furthermore, despite the proliferation of speculation about different category-learning theories in the domain of phonetics, rigorous experimental tests of rival claims have not often been attempted.

Along these lines, I created an experiment in which people learned German speech segment categories. In particular, they learned categories of fricatives—speech sounds created by the partial obstruction of the airflow in the vocal tract such that it creates audible turbulence, or frication—that are used in German but not in English, the voiceless palatal fricative ([ç], found in *ich*, ‘I’) and the voiceless velar fricative ([x], found in *ach*, ‘but’). The fricatives were associated with various colored squares, with the categories being learned changing from condition to condition based on which sounds are paired with which squares. This allows for a very flexible design. Below, I outline the results from two populations of interest when learning different speech sound categories: native speakers of American English, who are relatively inexperienced with these fricatives, and native German speakers, who have had decades of experience.

2.2.1 Experiment 1: Learning German Fricative Categories

I was interested in how easy it would be for participants to learn categories within a continuum of fricative sounds in order to compare the theories of category learning outlined in the chapter above. Learning categories of German fricatives is largely a departure from the prior linguistic experiences of American English speakers. American English speakers generally do not have familiarity with such sounds in a productive linguistic context. Although the velar fricative is present in some varieties of Scottish English and in some of the more common second languages spoken in the US (e.g., Spanish, Hebrew), the palatal fricative is much rarer, making the need to distinguish between the velar and palatal fricative to be an uncommon one for English speakers.

That said, it is unlikely that the American English speakers were *entirely* naïve to the sounds that they were being trained on. In fact, many of the native English-speaking listeners had learned languages that have the voiceless velar fricative as a part of their consonant inventory, such as Spanish or Hebrew, in an educational context. Although this likely does not represent a strong input, it is possible that such training would influence the acquisition of categories in some respect. One way to address such critiques is to examine a group of learners who would certainly have experience with the sounds in question: native German speakers. Unlike English speakers, German speakers' experience with the palatal and velar fricatives used here is, without question, meaningful, deep, and thorough. If German speakers show the same patterns of learning, it would suggest that any findings in the present experiment cannot be the result of

English speakers' sporadic exposure to the velar fricative alone. On the other hand, any differences between German and English native speakers could reveal the effects of expertise on category learning with these stimuli.

2.2.1.1 Participants

68 participants were recruited at the University of Maryland, College Park. Participants were compensated either for class credit in introductory Linguistics or Hearing and Speech Sciences department classes or with financial compensation. Data was excluded from 3 participants who had accrued more than incidental exposure to the German language, either through formal training or by living in a German-speaking country for at least a month; from 1 participant who was missing a demographics sheet; from 6 who were out of the target age range; and from 1 whose data file was corrupted. The participants remaining (*n* = 57) came from a typical undergraduate population (age *M* = 20.2, Range = 18-27, 34 female, 17 male, 6 not stated). All participants self-reported normal hearing and no history of speech or language disorder.

63 native German-speaking participants were recruited at the University of Tübingen. Participants were given €5 as payment for their participation in the task, and were generally recruited from linguistics-related listservs on campus or from previous participation in experiments within the linguistics department at the University of Tübingen. Two participants were excluded due to technical issues during the experiment, leaving a total of 61 participants. Of the 61 participants remaining—all young adults, aged between 19 and 34 (*M* = 23.7)—9 were male, 51 were female, and 1 did not indicate a gender. All participants gave informed consent, which was conducted in German. Data

collection was performed in line with German ethics standards, which do not require explicit ethics panel review for language-related experiments.

2.2.1.2 Materials

Both participant groups learned categories within a continuum of voiceless fricatives ranging from a voiceless palatal fricative [ç] to a voiceless velar fricative [x]. To create the stimuli, materials from a previous study (Key, 2014) were used as a starting point for this continuum. When given to me, the [x] and [ç] endpoints of the palatal-to-velar continuum had been excised from tokens produced by a native speaker of German, selected from a variety of recordings of [ç] and [x] in nonword frames. The tokens were judged to be representative of each category according to diagnostic acoustic features. The now-isolated tokens were cut at zero-crossings, with the longer token cut in size to match the length of the shorter token, and the peak intensities of each file were scaled to an identical 0.9 Pa. I then linearly combined the spectral content of these natural tokens using Praat (Boersma & Weenink, 2001) to create a 10-step continuum, with intermediate points that entailed a linear combination of the acoustic noise that characterizes each fricative. The steps were numbered from the palatal end of the continuum, with step 1 defined as the most palatal item and step 10 as the most velar item, with each intermediate number indicating the precise titration of the two endpoints.

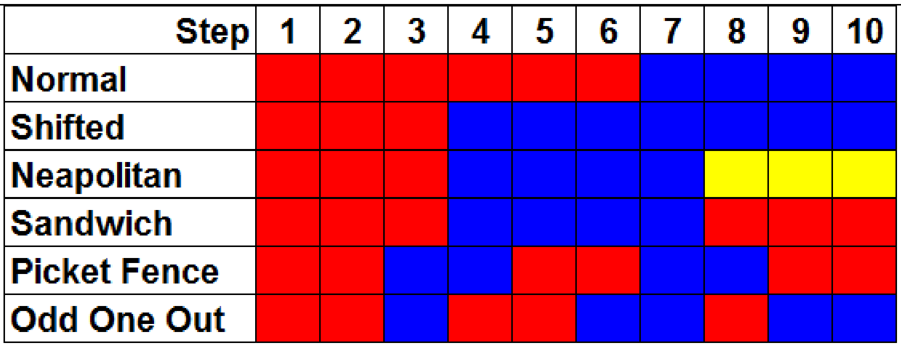
In using these materials, it is clear that a level of validity was sacrificed. Flat distributions with perfectly covarying cues are not typical for speech sound categories, particularly ones with only 10 items. Studies of cue trading, for example, have shown that many, if not all, phonetic contrasts are signaled with a wide variety of cues, all capable of combining together in many different ways to yield a coherent percept (Repp, 1982). The rich tradeoffs between these cues were not present in the present dataset. Additionally, of course, there is naturally more variability in the categories shown than can be accommodated with a flat distribution.

In this case, however, linear combination means that whatever multiple cues that listeners use to perceive the differences in place between these two fricatives are completely and inextricably correlated. This continuum therefore provides an avenue to measure the perception and acquisition of simple phonetic categories, akin to unidimensional voice onset time (VOT) continua used to examine the perception of word-initial voicing. Certainly, it is unclear how a model that cannot explain categorization in a unidimensional setting with a small number of items would scale up to more complex multidimensional phonetic categories with more realistic numbers.

2.2.1.3 Procedure

A time-to-criterion paradigm was used to explore learning using these items. Participants were first given brief instructions, telling them that they would hear speech sounds and that they would be asked to pair them with colored squares using the keyboard in front of them. They then heard a speech sound, 95ms long, from the 10-step continuum. This sound was presented simultaneously with three colored squares: blue, yellow, and red. Participants were given five seconds to pair the sound that they just heard with a square using one of three buttons on keyboard. They then received feedback about their selection in line with the condition they had been assigned to, as described below, which appeared 250ms after the participant selected a square and stayed on the screen for one second. The feedback took the form of a yellow “X” if the participant responded incorrectly or a green check mark if the participant responded correctly. The feedback was followed by a 500ms inter-stimulus interval (ISI). The order of trials was randomized in blocks of 10 steps each, such that participants heard all 10 steps every 10 trials (although with no predictable intra-block order). The distribution of trials was thus uniform, on average, across the experiment. Participants heard trials until one of two conditions was met: either 450 trials elapsed, or when the participant responded correctly to 90% of the last quasi-block (the last 10 unique items), which could span portions of the last two successive blocks. This meant that participants had to correctly respond to a wide swath of items along the continuum in order to complete the experiment early.

There were six conditions, assigned on a between-participant basis, with participant numbers in each condition approximately balanced for both participant groups. These conditions differed in which responses were considered correct on each trial. They are outlined below in Figure 5. Each row represents a single condition, with each column denoting a single step (ranging from the most palatal at left to the most velar at right). The boxes are colored in line with the correct response for each step in each continuum; for example, the correct answer for step 8 was yellow in the Neapolitan condition, red in the Sandwich condition, and blue in the Picket Fence condition. Note that there was no attempt to counterbalance item-color associations across participants, as I had no a priori reason to think that listeners should prefer to assign one end of the continuum to any particular color, as, say, might occur with synesthesia. Even for the five conditions in which only two responses were correct, all three possible responses were available.



Step	1	2	3	4	5	6	7	8	9	10
Normal	🟥	🟥	🟥	🟥	🟥	🟥	🟦	🟦	🟦	🟦
Shifted	🟥	🟥	🟥	🟦	🟦	🟦	🟦	🟦	🟦	🟦
Neapolitan	🟥	🟥	🟥	🟦	🟦	🟦	🟦	🟨	🟨	🟨
Sandwich	🟥	🟥	🟥	🟦	🟦	🟦	🟦	🟥	🟥	🟥
Picket Fence	🟥	🟥	🟦	🟦	🟥	🟥	🟦	🟦	🟥	🟥
Odd One Out	🟥	🟥	🟦	🟥	🟥	🟦	🟦	🟥	🟦	🟦

Figure 5. Experiment 1 conditions

The conditions differed in the numbers of possible categories and the number and composition of items assigned to each category. In the Normal condition, items were assigned to categories on the basis of the categorization preferences of English and German-speaking listeners from Key (2014), with a single boundary between items 6 and 7. In the Shifted condition, the category boundary was moved to between items 3 and 4. In the Neapolitan condition, the category boundary of the Shifted condition was preserved, while a third category, with a boundary between items 7 and 8, was added. In the Sandwich condition, the yellow stimuli from the Neapolitan condition were recoded as red, thus making the red category disjunctive (including items 1-3 and 8-10). In the Picket Fence condition, the assignment of items to categories went back and forth across the continuum, with items 1, 2, 5, 6, 9, and 10 assigned to red and 3, 4, 7, and 8 assigned to blue. Finally, in the Odd One Out condition, a boundary was placed between the red and blue categories between items 5 and 6 (near where the boundary was in the Normal condition), but with a single item on either side (items 3 and 8) being assigned to the category on the other side of the boundary (blue and red, respectively).

Almost all theories of category learning make identical predictions about four of these categories, albeit for varied reasons. Both the Picket Fence and Odd One Out conditions should be challenging. It is very hard to imagine a simple decision bound that can describe the category-learning parameters in both conditions, which each have many category boundaries. Prototype theories would posit completely overlapping prototypes in the Picket Fence condition, given its symmetry, while in the Odd One Out condition the odd one out would be very close to the center of each condition's prototype, meaning that that item should be frequently misclassified. To the extent that the items are hard to tell apart from each other, an exemplar-based learning account would suggest that the conditions would lead to a great deal of guessing behavior on the part of the learners. Given that these category structures are hard to learn for both exemplar and decision-bound theories, both systems in dual-systems models should struggle with learning.

Meanwhile, two of these conditions should be quite simple to learn at least for English speakers if one assumes that listeners come in with no priors about the categories present in this dataset: the Normal and Shifted conditions. Decision-bound theories predict that the single boundary in each condition should be easy to detect. Prototype theories would assign a prototype exactly in the middle of each category distribution. Exemplar theories would have no problems distinguishing between the items in each category. And both systems in dual-system models could easily learn the categories. However, in this case, it was of interest whether the biases present without training in Key's (2014) experiment would persist in the present dataset, particularly for the native German speakers, who had extensive experience with this continuum. This made the distinction between the Normal and Shifted conditions a noteworthy one.

The most interesting conditions were the remaining two, the Neapolitan and Sandwich conditions. Here, the theories make divergent predictions. Under prototype theories of category learning, the Neapolitan condition should be easy, while the

Sandwich condition should be hard. For the Neapolitan condition, prototypes would be found in the middle of each category, allowing for easy categorization of new items; in the Sandwich condition, meanwhile, the symmetrical nature of both categories would lead prototypes to be placed in the middle of the continuum for *both* categories, making learning challenging. Under exemplar-only theories of category learning, meanwhile, both categories should be equivalently easy to learn. The only difference between the conditions is in how stimulus steps 8 through 10 are categorized; as all else (including confusability between items) is held constant, there should be no difference in learning. The behavior of decision-bound models, meanwhile, depends on the treatment of the disjunctive red category according to these models. This is a matter of some debate, covered in a great deal more detail in the discussion of the present experiment. Some models do suggest that the disjunctive categories should be harder to learn than the non-disjunctive Normal and Shifted categories. The behavior of dual-system models depends on the decision-bound model that one adopts. If disjunctive categories are harder to learn under the decision-bound criteria used in the model, this might make the categories harder to learn until control of category learning passes to the similarity-based system.

2.2.1.4 Analysis

Most analyses for category-learning studies include a metric of the proportion of trials correct over time, averaged across blocks. Such experiments are based on multiple blocks, perhaps spaced across many sessions, with participants never quite approaching an optimal learning strategy (Nosofsky, 1986). In the present study, most participants successfully learned categories within the allotted 450 trials, making averaging within blocks problematic.

One crude but surprisingly effective measure to compare conditions is to bin participants into one of three groups, based on how quickly they learned to pair the speech sounds to the colored squares. In this case, learning was defined as giving a correct answer on 9 out of the most recent token of each of the 10 different items. The time to this learning criterion could therefore range from 9 trials (if a participant answered the first nine trials of the experiment without a single error) to 450 trials. "Fast learners" learned in less than 225 trials. "Slow learners" took more than 225 trials to learn the pairings of speech sounds to colored squares. "Non-learners" did not learn the pairings of speech sounds to colored squares in the 450 trials given to them. The number of fast learners, slow learners, and non-learners could then be compared across conditions.

A second analysis stream made use of generalized linear mixed models, as instantiated in the lme4 package in R (Bates, Maechler, Bolker, & Walker, 2016). Using mixed models meant that trial-by-trial patterns of learning over time could be broken down by condition, after factoring out differences in learning between individuals and between items. Performance on each trial was coded as a binary response variable, with correct trials coded 1 and incorrect trials coded 0. Condition (possible values Normal, Shifted, Neapolitan, Sandwich, Picket Fence, and Odd One Out, with Neapolitan as the reference level) and native language (English or German, with English as the reference level) was coded as a factor, while trial number was coded as a continuous variable (and rescaled to have values between -0.5 and 0.5 to aid in model convergence). I included fixed effects of condition, trial number, and native language, as well as the interactions between them (both two-way and three-way), random intercepts by participant and by item, and random slopes for trial number by participant (i.e., participants could vary in their individual learning abilities), and random slopes for condition by item (i.e., items could vary in how difficult they were between conditions). As in previous studies (Chandrasekaran, Yi, et al., 2014; Scharinger, Henry, & Obleser, 2013), the interaction between trial number and condition was used as a proxy for different learning rates across conditions.

This model was then used to compare performance between conditions using the lsmeans package (Lenth, 2016), which allows for Tukey-adjusted comparison of least-squares means between factor levels. In practical terms, this allows for post hoc comparisons of learning across each of the conditions, with the goal of illuminating meaningful contrasts highlighted above: between the Normal and Shifted conditions (to determine the power of the pre-existing bias towards categories split between steps 6 and 7), the Normal and Odd One Out conditions (to check that the “easy” and “hard” conditions were truly separated from one another, and to what extent), and the Neapolitan and Sandwich conditions (to evaluate the split between disjunctive and non-disjunctive categories), as well as contrasts between participant groups.

2.2.1.5 Results

The results are presented in Figure 6. Each section represents a single condition (labeled at left), divided into German- and English-speaking participants, with individual participants shown as a single circle within each combination of condition and native language. Participants’ horizontal displacement along the graph shows the number of trials needed for that individual participant to learn the correct pairings of speech sounds to colored squares for that condition, with participants clustered along the vertical line at far right being participants who failed to learn within 450 trials. Vertical displacement from each condition line is meant to highlight the number of individual participants at each location and is not of itself informative. Median trials to learn for each combination of participant group and condition is shown with a red “X”. As can be seen from the graph, there is a stark difference between the first three conditions and the last three conditions. Participants generally found the Normal, Shifted, and Neapolitan conditions much easier than the Sandwich, Picket Fence, and Odd One Out conditions. By and large, participants were not strongly biased towards any particular category cross-over point, as they found the Normal and Shifted conditions approximately equally easy to learn. With regard to the effects of the participants’ language background, the participant groups resembled each other in their performance on this task. The categories that German speakers found relatively difficult were also the ones that English speakers found relatively difficult; conversely, the categories that German speakers found easier were also ones that English speakers had found easier.

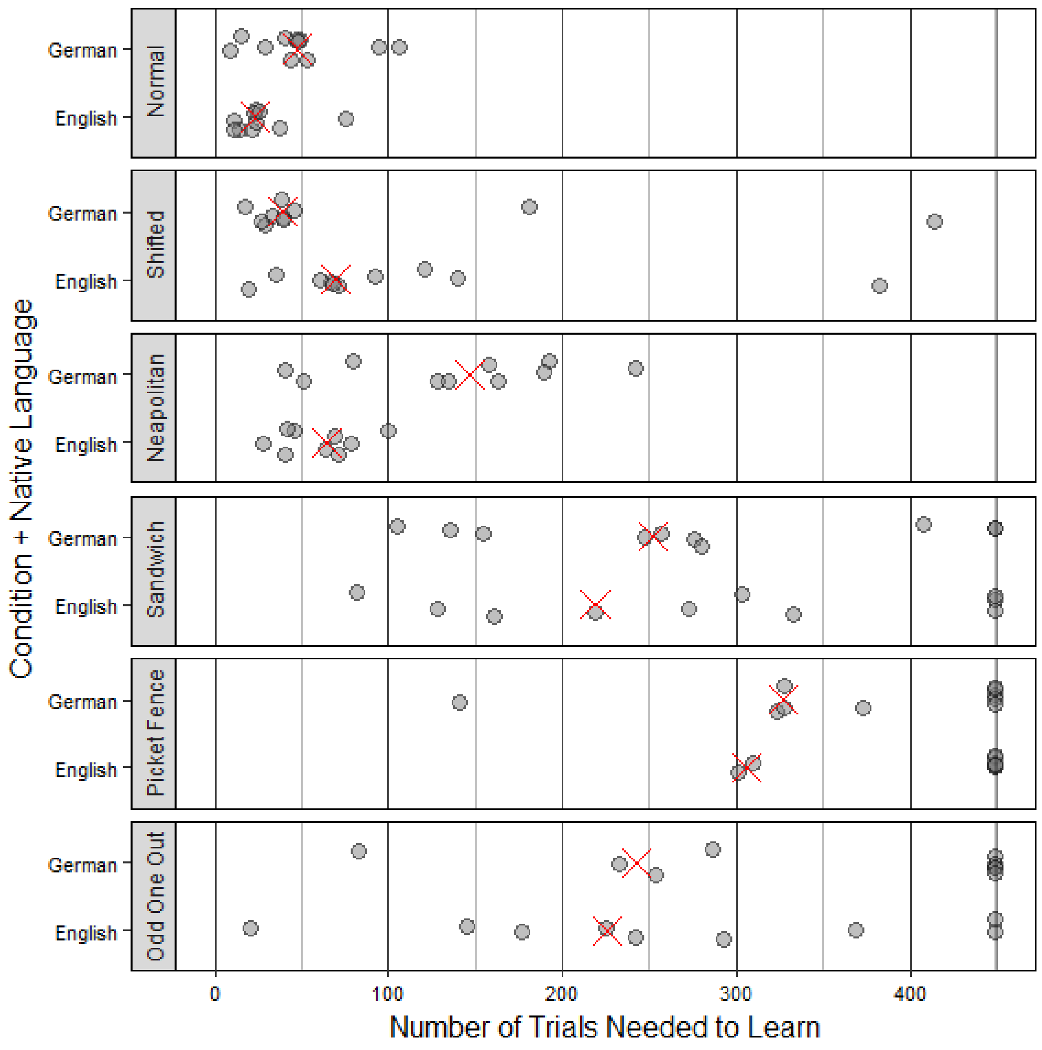


Figure 6. Experiment 1 results by participant

Clearly, of course, these results (as well as those of Chapter 2.3) do require a caveat related to the sample size. A total of roughly 60 native-English-speaking and native-German-speaking participants implies approximately 20 participants total for each condition. This is not a very large sample size, especially given the hypothesized presence of null effects between conditions. At least some future studies will be run online, to help in participant recruitment, with all participants required to complete all 450 trials (in order to provide more data across all participants). Still, even with the data that was collected, some interesting inferences can be drawn. One way to examine this dataset is to compare median times to learn across participant groups and conditions, and to divide participants into fast learners, slow learners, and non-learners. Table 1 shows the outcome of these divisions, measured in terms of median time to learn (MTTL) for participants who learned within 450 trials, and by participant group. Fast learners learned within 225 trials; slow learners took between 225 and 450 trials; and non-learners had not learned by the end of the experiment. Note that the median times to learn (MTTL) within the table sometimes include a very small number of participants in certain combinations of participant groups and conditions (e.g., just two native English speakers in the Picket Fence condition); these values are shown merely for descriptive purposes.

Condition	**Language**	**MTTL**	**Fast Learners**	**Slow Learners**	**Non-Learners**
Normal	English	23.5	10	0	0
	German	47.0	11	0	0
Shifted	English	69.5	9	1	0
	German	39.0	9	1	0
Neapolitan	English	64.0	9	0	0
	German	146.5	9	1	0
Sandwich	English	219.0	4	3	3
	German	252.5	3	5	2
Picket Fence	English	306.0	0	2	7
	German	327.0	1	4	5
Odd One Out	English	226.0	3	4	2
	German	243.0	1	3	6

Table 1. Summary table for Experiment 1

A series of rough approximations were made to examine differences in time to learn across conditions. First, the conditions were broken into two groups, one including the Normal, Shifted, and Neapolitan conditions, and the other including the Sandwich, Picket Fence, and Odd One Out conditions. The first group contained conditions without disjunctive categories; the second group contained conditions with disjunctive ones.

Next, individual participants were split into two groups, based on their time-to-criterion: fast learners, and a combined group made up of slow learners or non-learners. This forms a 2 × 2 contingency table.

Using Fisher's exact test, it is readily apparent that English-speaking participants in the two condition groups differ in their likelihood to be a fast learner versus a slow learner or non-learner, $p_{\underline{\text{two-tailed}}} = 1.04 \times 10^{-8}$. This is also true if one compares the likelihood to be a fast or slow learner versus being a non-learner, $p_{\text{two-tailed}} = 4.30 \times 10^{-5}$. Indeed, regardless of where one puts the boundaries between fast learners, slow learners, and non-learners, there is no boundary condition such that there are more "faster learners" in the second set of category conditions than in the first set. Just like with the English-speaking participants, the German-speaking participants were much more likely to be fast learners, or to learn at all, in the continuous category conditions than in the discontinuous category conditions. For example, performing a Fisher's exact test on participant counts when crossing category continuity with speed of learning (pitting fast learners against slow learners and non-learners) yields a p value of 4.57×10^{-10}, while doing the same while breaking speed of learning up into learners (fast or slow) and non-learners yields a p value of 7.67×10^{-5}.

These qualitative findings, however, are not the whole story. Figure 7 shows Loess-smoothed plots of performance over time for the English (solid lines) and German (dashed lines) speakers in each condition. Performance on the part of participants who finished early (before 450 trials) is interpolated with 90% accuracy in the graph above, leading to the flat level of performance in especially the Neapolitan and Normal conditions in the figure below.

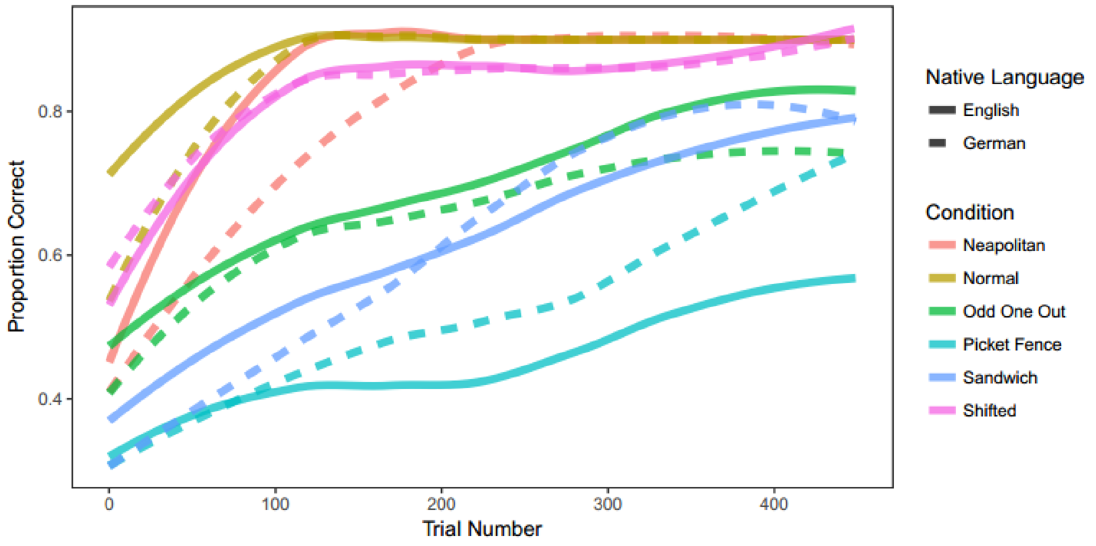


Figure 7. Experiment 1 results, Loess-smoothed

A detailed exploration of the results was undertaken using mixed models, through a combination of model comparison and post doc comparisons using the lsmeans package (Lenth, 2016). First, the intermediate model described above (see model specifications in Table 2) was compared to models that lacked various combinations of the fixed effects and interactions in question. The fixed effects of Condition use Neapolitan as the reference condition, while the fixed effects of native language use English as the reference condition. The intermediate model fit better than a model that lacked fixed effects of condition and its interactions, $\chi^2(20) = 115$, $p < .001$; better than a model that lacked fixed effects of trial number and its interactions, $\chi^2(12) = 163$, $p < .001$; and better than a model that lacked fixed effects of native language and its interactions, $\chi^2(12) = 31.8$, $p = .001$. In other words, the condition a participant was assigned to and the native language of a participant both affected learning rates, and participants did successfully learn over time.

Fixed Effect	***β***	**SE**	***z***	***p***

Intercept	5.68	1.88	3.01	.003
Condition: Normal	0.180	3.35	.054	.96
Condition: Odd One Out	-5.27	1.99	-2.65	.008
Condition: Picket Fence	-5.96	1.89	-3.16	.002
Condition: Sandwich	-5.49	1.88	-2.92	.004
Condition: Shifted	-4.21	1.91	-2.21	.03
Trial Number	12.3	4.25	2.91	.004
NL: G	-3.88	1.91	-2.04	.04
TN × C: N	-1.44	7.33	-0.20	.84
TN × C: OOO	-11.3	4.29	-2.64	.008
TN × C: PF	-11.7	4.26	-2.75	.006
TN × C: Sa	-11.2	4.27	-2.63	.009
TN × C: Sh	-10.4	4.31	-2.40	.02
NL: G × C: N	4.46	3.64	1.23	.22
NL: G × C: OOO	4.14	1.94	2.14	.03
NL: G × C: PF	4.04	1.92	2.10	.04
NL: G × C: Sa	3.94	1.93	2.04	.04
NL: G × C: Sh	4.14	1.94	2.14	.03
TN × NL: G	-8.13	4.38	-1.85	.06
NL: G × C: N × TN	10.7	8.07	1.33	.18
NL: G × C: OOO × TN	8.69	4.46	1.95	.05
NL: G × C: PF × TN	8.57	4.41	1.94	.05
NL: G × C: Sa × TN	9.02	4.43	2.04	.04
NL: G × C: Sh × TN	7.72	4.48	1.73	.08

Table 2. The best-fitting model for Experiment 1

Interactions with native language also had a significant impact on model fit. Comparing the intermediate model to one lacking the three-way interaction between condition, native language, and trial number and the two-way interaction between condition and native language to the intermediate model also yielded a significant decrease in model fit without the interactions, $\chi^2(10) = 23.5$, $p = .009$. Some conditions had higher (or lower) baseline rates of learning for German speakers than English speakers. Subtracting the two-way interaction between trial number and native language (and the three-way interaction) also hurt model fit, $\chi^2(6) = 19.4$, $p = .004$. German speakers learned (slightly) slower than English speakers across conditions. Even just removing the three-way interaction also decreased model fit, $\chi^2(5) = 17.4$, $p = .004$. German speakers were particularly slow to learn the categories in certain conditions.

Post-hoc comparisons were used to determine some contrasts of interest. In particular, I was interested in whether any of the conditions differed in their difficulty between native speakers of English and native speakers of German. I was also interested in contrasts between key conditions: between the Normal and Shifted conditions, Normal and Odd One Out conditions, and Neapolitan and Sandwich conditions. In total, then, there were 12 contrasts of interest; the multivariate *t* adjustment was used to ensure that *p* values were adjusted appropriately for the number of comparisons.

2.2.1.5.1 Normal vs. Shifted

There was no significant difference for English speakers in the trial number effect between the Normal condition and the Shifted condition, $z = 1.81$, $p = .49$. It was no harder for the English-speaking participants to learn the Normal condition than the Shifted condition (or vice-versa). For the German speakers, meanwhile, there was a significant difference, $z = 4.60$, $p < .001$, with the Normal condition being significantly easier than the Shifted one. This suggests that the bias that both English-speaking and German-speaking listeners demonstrated in Key's (2014) exploration of German fricative contrasts was quite labile for English speakers, but more intractable for German speakers, as might be expected by the relative amounts of experience each group had. At the same time, these results may have been driven primarily by the behavior of the two German-speaking outliers in the Shifted condition.

2.2.1.5.2 Normal vs. Odd One Out

Surprisingly, there was no significant difference for English speakers between the Normal and Odd One Out conditions in learning, $z = 2.31$, $p = .18$, despite having very different median times to learn. This is unexpected under any category-learning theory. However, this may be due in part to the very small number of trials that most participants in the Normal condition needed to learn the pairings of sounds to colored squares or the variability present in learning within the Odd One Out condition. This, in turn, increased the standard error estimates in the model for both conditions. The model simply may not have been provided with sufficient data to uncover a significant difference. The effect was significant for German speakers, $z = 4.80$, $p < .001$, implying that German speakers were hobbled by the presence of the outlier category members when learning these categories.

2.2.1.5.3 Neapolitan vs. Sandwich

The final comparison undertaken was between the Neapolitan and Sandwich conditions. The prototype and some decision-bound theories of category learning on the one hand predict there to be a significant difference between these conditions; exemplar-only and other decision-bound theories, on the other hand, largely predict that learning in these two conditions should be roughly the same. In the end, a significant difference was found in learning between these two conditions for English speakers, $z = 3.02$, $p = .02$; the Neapolitan condition was much easier to learn than the Sandwich condition. This was also true for German speakers, $z = 4.12$, $p < .001$. This contrasts with the predictions of exemplar-only and certain decision-bound theories of learning.

2.2.1.5.4 English vs. German Participants

Given the similarities between the English and German speakers in the key contrasts, what drove the significant differences between the participant groups, then? Visual inspection appears to show the Neapolitan condition being harder for German speakers than English ones. However, direct comparisons showed no significant differences between the participant groups for any condition. The Normal (*z* = -0.123, *p* = 1), Shifted (*z* = 0.719, *p* = 1), Neapolitan (*z* = 1.34, *p* = .29), Picket Fence (-0.0964, *z* = -0.499, *p* = 1), Sandwich (*z* = 0.328, *p* = 1), and Odd One Out (*z* = -0.859, *p* = .99) conditions showed no significant differences between the English speakers and the German speakers in their success over time.

2.2.1.5.5 Anti-Disjunctivity: Further Findings

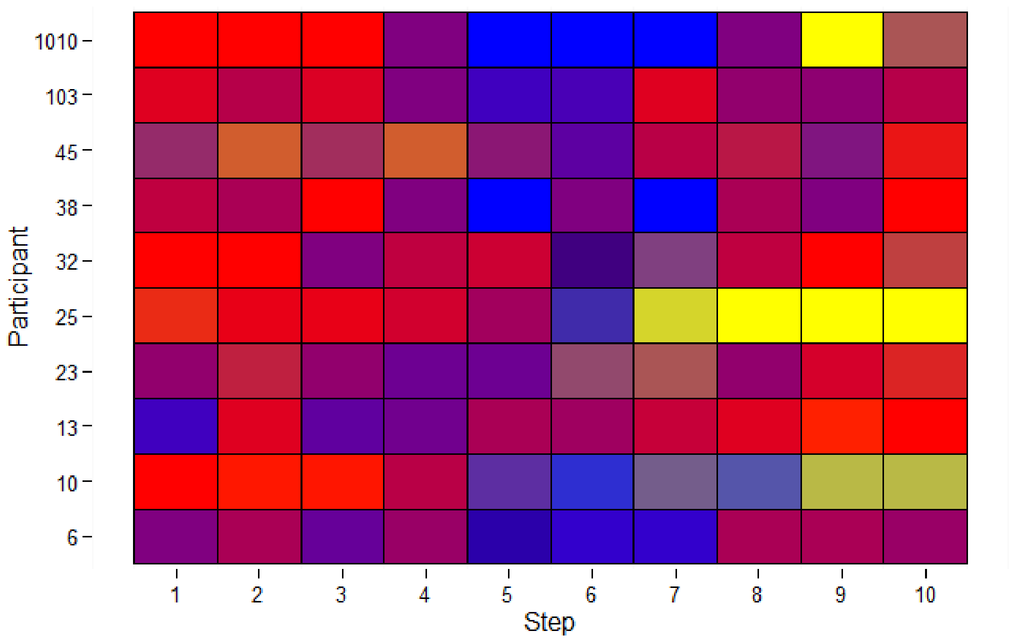


Figure 8. Experiment 1 Sandwich final section results

Here, the behavior of non-learner participants seems particularly instructive. Figure 8 shows the responses for the participants in the Sandwich condition within the last 25% of trials in the experiment. Each row corresponds to the behavior of a single participant, with 10 columns that correspond to the responses participants gave when they heard those last trials. Each cell is coded in line with the proportion of responses from each category. Cells that are entirely blue, red, and yellow indicate that participants responded 100% with that color category for that step within the last 10% of trials administered, while cells with intermediate colors represent combinations of responses. For example, purple cells represent some red and some blue responses for those steps towards the end of the experiment. If participants are basing their responses only on exemplars, the ends of the continuum should be reddish, and the middle stimulus steps should be bluish, with some level of purple (reflecting both blue and red responses for a certain point along the continuum) being likely around stimulus boundaries. This is what is seen for participants 6, 38, and 103, for example.

Yet there are also participants with quite different patterns of responses. Participant 13 had an almost linear grading from uniformly red on the velar end of the continuum to uniformly blue on the palatal end of the continuum. Most strikingly, participants 10, 25, and 1010 continued pressing “yellow” for the velar end of the continuum until almost the very end. They were so certain that there must be three categories within the continuum that they are giving a yellow response through to the end of the experiment even despite the fact that they are always told that such a response is wrong! The responses of participant 25, in particular, show a clear categorical separation: Steps 1-5 as red, 7-10 as yellow, and 6 as blue, with some noise in responses. This largely contrasts with the predictions of exemplar-only accounts of category learning.

2.2.2 Discussion

Below, I flag two issues of particular interest that the results above highlight: first, the acquisition of disjunctive categories, and, second, the role of expertise in category learning.

2.2.2.1 Disjunctive Categories in Exemplar-Only Models

These results are challenging to accommodate using exemplar-based phonetic learning models, but for reasons that are somewhat counterintuitive: exemplar models are too *good* at fitting this dataset. As discussed in the literature review for category learning, exemplar models are excellent learners; they are capable of learning any variety of categories, and, over time, they predict that categorization should be optimal (Ashby &

Alfonso-Reese, 1995). Put another way, exemplar-only models "basically [predict] that given enough experience with training exemplars, participants' response patterns should eventually approximate the underlying category descriptions" (McKinley & Nosofsky, 1995, p. 145). However, this clearly not always the case, as participants often show suboptimal behavior when learning new categories (Ashby et al., 2001). That was the case here for the Sandwich, Odd One Out, and Picket Fence conditions. Although the poor performance in the Odd One Out and Picket Fence conditions is explicable within an exemplar framework, it is challenging to incorporate the distinction between the Sandwich and Neapolitan categories using exemplar-only models.

To explore why, consider the findings in light of the Generalized Context Model (GCM) of Nosofsky (1986), which is one of the most influential of the single-system exemplar-based models. Nosofsky proposed that categorization was essentially reducible in its core to identification. Sorting a new item into a category was the process of assessing the similarity of that new item to the previous items observed, then assigning the new item a category label based on the category labels of the most similar previous items. The computational implementation of the GCM is fairly simple. The distance between a current item and previous ones is computed using a Gaussian distance function. This distance function is used to compute the weight that each item has towards categorizing the stimulus into any of the possible categories under consideration. The category with the greatest summed weight then "wins", and the new item is assigned to that category.

In no case, then, does the category label of any particular item affect the difficulty of categorization. The only determinant of difficulty is *discriminability*. From exemplar-only views of phonetics, “the strength of the representation at a location on the map depends merely on the number and recency of the exemplars at that location” (Pierrehumbert, 2003, p. 132). The more tokens, and the more recent the tokens, the stronger the discriminability. If items are hard to discriminate, it will be hard to learn categories, as items that are far away from a new token are mistakenly brought to bear on its categorization. If items are easy to discriminate, new tokens will be labeled accurately (or, at least, in line with the immediately adjacent exemplars). This provides a perfectly cogent reason for why the Odd One Out and Picket Fence conditions were hard to learn. If items were not very discriminable, tokens from the “wrong” category would be erroneously sampled and contribute to incorrect categorization. It does not take a particularly wide sampling of category tokens in either condition to pick up some wrong tokens.

However, where the mechanisms of the GCM go afoul is when comparing the Neapolitan and Sandwich conditions. The Neapolitan condition was quite easy for the English-speaking participants, while it was harder for the German-speaking ones. The Sandwich condition was more challenging for both participant groups. This occurred despite the fact that the boundaries between the categories were *exactly the same* in the two conditions. That is, no matter where a novel item fell within the speech sound continuum, the distances to adjacent and non-adjacent categories were identical across the conditions; discriminability was not different between conditions. The only thing that changed was the category label of one side of the continuum, and that was enough to entirely alter participants’ times to learn. This was not merely a byproduct of the fact that participants in the Sandwich condition had to learn to ignore a possible response.

Participants in the Normal and Shifted conditions had to do the same thing, but generally had little problem learning the categories of colored squares.

A more subtle point of note relates to the performance of participants who failed to learn the pairings of speech sounds to colored squares. According to the GCM, there are essentially two possible outcomes for participants in conditions such as the Picket Fence condition: learning or guessing. Participants will learn the categories under consideration if they are capable of discriminating nearby items on the continuum. If they are not capable of discriminating adjacent items, however, participants will instead sample from a wider distribution of items. In a condition such as Picket Fence, sampling from a wide distribution across the continuum leads to essentially equal numbers of blue and red category items being considered for each position along the continuum. Participants would therefore be approximately at chance in their responses. But this is not the behavior that was observed even for participants who failed entirely. Instead, many participants in the Picket Fence, Sandwich, and Odd One Out conditions showed response patterns that differed systematically from chance, as can be seen for the Sandwich condition in Figure 8.

2.2.2.2 Disjunctive Categories in Decision-Bound/Multiple-System Models

Other models of category learning seem to hold more promise. One alternative to a single-system approach is to incorporate multiple learning systems into the model (Chandrasekaran, Yi, et al., 2014; Maddox & Chandrasekaran, 2014). Under these theories, an explicit, rule-based system precedes an implicit, procedural-learning-based system. This has the possibility of reversing many contemporary theories of phonetic category learning, where abstraction is generally said to follow and require memorization of specific instances (Maye, Weiss, & Aslin, 2008; Maye, Werker, & Gerken, 2002; McMurray, Aslin, & Toscano, 2009), but has more in common with more recent approaches to phonetic category learning that incorporate neurobiological insights (E. B. Myers, 2014). One of the benefits of a dual-system approach is that the findings spurred by exemplar-based approaches to phonetic category learning do not need to be discarded. If the explicit learning system shapes the tectonic plates that determine the rough alignment of categories to each other, implicit learning forms the processes of erosion and deposition that determine the precise contours of the category topography. Both are necessary for a complete understanding of the human capacity for phonetic category learning.

On the face of it, then, dual-system theories provide a ready way to explain the distinction found between the Neapolitan and Sandwich conditions: the categories in the Neapolitan condition are learned through the fast-acting rule-based system, while the categories in the Sandwich condition are learned through the slower-acting similarity-based system. This is complicated, though, by the characteristics of the rule-based system under many dual-system theories. A priori, there is no particular reason why disjunctive rules (e.g., a category item is red if it is below step 4 *or* above step 7) would not be just as easy to learn as non-disjunctive rules. Many dual-system proponents have suggested that disjunctive categories may sometimes be learned using the *rule-based* learning system, rather than the similarity-based one (Minda, Desroches, & Church, 2008; Zeithamova & Maddox, 2006). In one instance of this, for example, categories were defined in terms of the pitch of a non-linguistic tone and the orientation of a visual Gabor patch. One category was defined as items either when the tone was high-pitched and the orientation of the patch was vertical or when the tone was low-pitched and the orientation of the patch was horizontal. This was treated as an example of rule-based disjunctive categorization because the dimensionality of the stimuli was such that information did not need to be integrated across the visual and auditory modalities; learners could combine separate information from sound and vision (Maddox, Filoteo, Hejl, & Ing, 2004). As such, strategies on the part of learners that the authors described as "rule-based" led to approximately the same level of performance as strategies that could be described as "similarity-based" even though the categories that were being learned were clearly disjunctive.

The dual-system theories discussed above, however, are not the only possibilities outside the realm of exemplar-only theories of category learning. The RULes and EXceptions model (RULEX) provides another avenue of exploration (Nosofsky et al., 1994). In RULEX, categories are formed through a multi-stage process. First, learners try to identify simple rules to characterize the categories being taught to them. If those rules are categorically (or close-to-categorically) successful at characterizing the stimulus space, the rules are kept unaltered. If they are entirely unsuccessful, the learners try instead to learn more complex rules (i.e., multidimensional ones) that involve interactions between categories. And if the rules are moderately successful—say, successful 75% of the time—learners memorize exceptions to the rule, with the number and specificity of the exemplars depending on the learner's memory constraints.

Although RULEX was initially applied to a learning situation with just two categories, using items that varied in a binary fashion across four dimensions, its continuous-dimension update (Nosofsky & Palmeri, 1998) applies it to continuous dimensions such as the ones explored here. Still, although a model that uses rules and memorized exceptions to those rules could be useful to help describe this situation, even the continuous RULEX suffers from some key deficits. For example, a model that could quickly and easily memorize exceptions for mostly-valid rule-based characterizations of categories would struggle to explain why the Odd One Out condition was so challenging for participants. More dauntingly, though, the continuous RULEX also relies on pre-specification of possible rule/exception pairings by hand.

One way to incorporate some of the insights of RULEX into a computational framework is to take a Bayesian approach to category learning, an approach that is growing more and more common within cognitive science (Jacobs & Kruschke, 2011). Such ideas help make up the Rational Rules model of concept learning (Goodman, Tenenbaum, Feldman, & Griffiths, 2008). In the Rational Rules system, hypotheses take the form of rules. In learning scenarios that include non-disjunctive dimensions, these rules are formed from conjunctions or disjunctions of sets that describe parts of a particular dimension. Participants make responses in line with the small number of hypotheses that they are entertaining at any one particular point about the categories that they learn, with a small probability of responding incorrectly. Individual items also have the chance of being labeled as an outlier if they belong to a category unexpected by the rules currently under consideration. Simple rules are preferred to more-complicated ones due to a strong prior for simple rules. Under Rational Rules, participants have strong priors towards simple categories, just like the ones here. Learning more complex rules, including ones that require disjunctions or conjunctions, takes time. Indeed, such stipulations could be integrated into the rule-based system of dual-system models based on COVIS (Ashby et al., 1998) to produce similar biases against disjunctive rules.

In sum, then, whether disjunctive categories should be any harder to learn than non-disjunctive ones under dual-system category models depends on the properties of the rule-based system in the model. If the rule-based system is biased against disjunctive categories, as in RULEX, the Sandwich condition becomes harder than the Neapolitan condition because listeners must fall back on the memorization of exceptions or similarity-based decision bounds in order to learn the disjunctive category of the Sandwich condition. If, on the other hand, the rule-based system is capable of learning disjunctive categories, there should be no reason for the qualitative split in learning times.

2.2.2.3 Expertise in Category Learning: German and English Speakers

Another key insight of this project is with regard to the importance of previous experience with relevant speech sound categories in phonetic category learning. As mentioned previously, the English-speaking participants in this study had only limited second-language exposure to the sound categories in question; although they may have perceptually assimilated the sounds to similar English categories, their experience with the tokens in question was likely minimal. The German speakers, meanwhile, had amassed decades of experience.

The idea of reshaping pre-existing speech sound categories is one that is usually tied to perceptual learning of speech. Perceptual learning, referring to changes in perception that stem from environmental input, has been most extensively studied in the visual perception literature (Goldstone, 1998). However, it has also been repeatedly and convincingly demonstrated in speech perception experiments (Samuel & Kraljic, 2009), with recently-played speech information determining later-occurring speech categorization. In many perceptual learning studies, participants are presented with sounds acoustically intermediate between two different segments: for example, intermediate between an [f] and an [s] (Eisner & McQueen, 2005; Norris, McQueen, & Cutler, 2003), between an [s] and an [ʃ] (Kraljic & Samuel, 2005), or between a [d] and a [t] (Kraljic & Samuel, 2006). These ambiguous sounds are placed in lexical contexts that disambiguate which category the sound belongs to. For example, a sound ambiguous between [t] and a [d] (denoted [?]) is more likely to be perceived as a [d] in the context of [ɑvəkɑ?o] ‘avocado’ but more likely to be perceived as a [t] in the context of [lunə?ɪk] ‘lunatic’ because neither ‘avocato’ nor ‘lunadic’ are words of English. Sufficient training leads listeners to treat the formerly ambiguous segments as unambiguous even in the context of non-biasing lexical items (Norris et al., 2003). Thus, when hearing a simple VCV non-word such as [a?a], listeners trained on [lunə?ɪk] will be more likely to hear [ata], while listeners trained on [ɑvəkɑ?o] will be more likely to hear [ada]. It is challenging for listeners to go back to hearing the items as ambiguous (Kraljic & Samuel, 2005) so long as the original talker is producing the test items as well (Eisner & McQueen, 2005). The perceptual learning that ensues can also be extended to similar but untrained lexical contrasts, such as other stop voicing contrasts for the [t] and [d] stimuli (Kraljic & Samuel, 2006).

Perceptual learning experiments generally only examine how training can shift the boundaries of different native categories. It is not clear to what extent perceptual learning studies can scale up to predict the acquisition of wholly (or even partially) novel categories within familiar phonetic spaces. Adult learners may be put into this situation when, say, learning three phonetic categories when their native language only has two. An example of this is with English-speaking adults learning three stop categories—aspirated, unaspirated, and voiced—in a continuum they are accustomed to having just two, aspirated and unaspirated (Beach, Burnham, & Kitamura, 2001; Pisoni, Aslin, Percy, & Hennessy, 1982). Adults come to these tasks with preconceptions about the sounds they have had exposure to; adults are, in many ways, language "experts". As such, the literature on category learning in experts may help inform the effects of native language expertise on speech sound learning.

Expertise leads to a variety of changes in learning non-linguistic categories. For example, chemistry students are increasingly likely to classify chemical reactions better in line with the underlying properties of the chemical reactions (for example, whether the reactions involved the creation of a precipitate) rather than some of their surface properties (for example, whether the reactions involved solids or liquids) as they gain experience in chemistry classes (Stains & Talanquer, 2008). Tree experts (e.g., landscapers, parks maintenance workers) are more likely than undergraduates to use ecological properties of tree types (say, the distribution of tree types, or the susceptibility of certain tree types to disease) rather than pure taxonomy to try to determine whether a disease found originally in one type of tree might affect a different type (Proffitt, Coley, & Medin, 2000). Similar results were obtained for commercial fishermen sorting fish into categories (Shafto & Coley, 2003).

These findings and others like them have led to a proliferation of approaches to explaining the changes seen in category learning with expertise. One review paper (Palmeri, Wong, & Gauthier, 2004) enumerated many of the theories of expertise in category learning. The accounts can roughly be sorted into one of two hypothesis classes. Under one class, experts are better categorizers because they are in some way better at using stored exemplars. This might be true because experts are using a similarity-based learning system more often (as in many dual-system models), because exemplars are better-tuned (i.e., new items lead fewer memories to be activated in determining categorization), or because perceptual noise is lower. Under the second class of hypotheses about expertise in categorization, meanwhile, it is the dimensionality of the space being learned that changes with experience. Examples of these changes include selectively attending to different perceptual dimensions in the category space; blurring distinctions between dimensions; and adding (or removing) entire dimensions by which categories can vary.

Under the class of hypotheses related to expertise in which expertise sharpens exemplar representations, learning should be better nearly across the board, particularly, under dual-system approaches, for categories that rely on a similarity-based system of learning. Why? For similarity-based approaches to learning, the primary determinant of how easy categories are to learn is how easily distinguishable individual tokens are from each other. New items are misclassified when speakers mistakenly activate tokens of the wrong category because either perceptual noise or a wide-tuned item activation could lead to errant activation of the wrong colored square category. Being “experts” under these conceptions of expertise makes this errant activation increasingly unlikely, thereby making it easier to correctly assign items to categories.

Under one of the set of theories in which expertise leads to reshaping of dimensionality in category learning, hypotheses are more challenging to construct. This is true because the dimensions over which speech sounds vary are manifold and poorly understood, meaning that changing those poorly understood categories becomes even harder to predict. Changes in dimensionality could have made the categories easier to learn, if, say, items assigned to the same category in the experiment were perceived to be more similar to each other due to shifts in dimensional focus. Alternatively, the categories could have been made more difficult because of perceptual warping of the acoustic space, as is hypothesized to underlie categorical perception according to some rational accounts of phonetic perception (Kronrod, Coppess, & Feldman, 2016).

Not many effects of expertise surfaced in the present study. Indeed, although the relative difficulty of the disjunctive conditions may have been different in some ways from those observed in English speakers, these differences were neither strong nor consistent nor significant when considered in isolation. This suggests that the idea that category expertise involves the sharpening of exemplar representations seems to miss the mark for the German speakers here. Under the GCM (Nosofsky, 1986), there is a scaling parameter that determines the range of exemplars that are activated during memory retrieval; sharpening this parameter is one way to actuate many of the exemplar-based effects of expertise described by Palmeri et al. (2004). Simulations that I have performed have indicated that changes in this scaling parameter can lead participants to learn the Picket Fence and Odd One Out conditions more quickly. Yet this is not the pattern observed; German-speaking participants, despite their orders of magnitude larger number of exemplars of each of these category items, do not behave any differently from English speakers in their mastery of the Picket Fence and Odd One Out conditions. Indeed, the fact that *none* of the conditions were significantly different depending on the participant group indicates that any perceptual warping that results from experience was relatively minor. English and German speakers seemed to perceive the stimulus continuum in similar ways.

2.2.2.4 Summary

To summarize, I trained English speakers and German speakers to categorize a continuum of German fricatives by having them pair the items along the continuum with colored squares, which were used as stand-ins for category labels. Each participant learned one of six conditions that differed in the assignment of items to categories. What I found was that some categories were harder to learn than others. Crucially, I found that disjunctive categories were generally harder to learn than non-disjunctive ones, even when the only difference between the categories was in the assignment of one set of items to a category label. This is difficult to model using exemplar-only theories of category learning, both inside and outside language, but easier to accommodate under prototype theories and some dual-system theories of category learning. English and German speakers were not very different from each other in their acquisition of categories. This suggests that improvements in perceptual sensitivity do not distinguish German-speaking experts from English-speaking novices in their acquisition of these categories.

However, these experiments do not directly get at the idea of modularity in speech perception. Although models of category learning and category expertise provide useful guides to understanding the results here, these findings were obtained only within the domain of speech perception. In the following chapter, I examine the acquisition of auditory categories outside the realm of language, in particular the acquisition of different musical instrument categories. Does the bias against disjunctivity hold even outside of language?

2.3 Non-Speech Category Learning

What, then, is the connection between phonetic category learning and category learning more generally? This question is a very relevant one to the question of whether speech is “special”. In studying this, I evaluate the claims of Liu and Holt (2011) that the processes of learning speech sound categories resemble the processes of gaining expertise with non-speech categories. At the same time, I use auditory materials, rather than, say, visual materials, to ensure that as much else as possible is held constant when assessing non-linguistic category learning, including modality. Judging the similarity of visual items to auditorily-presented linguistic ones would be challenging indeed. This comes with an additional benefit of probing auditory categories in greater detail; categories other than visual ones are very understudied in the non-linguistic category learning literature. To examine whether non-speech auditory category learning resembles phonetic category learning, I used non-linguistic categories as analogues to speech ones, with the goal of assessing whether non-linguistic categories are also subject to bias against disjunctive categories that was observed for both English and German speakers in the fricatives.

2.3.1 Non-Speech Materials

In order to examine the acquisition of rich and acoustically-complex non-linguistic categories, I created a continuum of synthetic musical instrument sounds. This was done using the Wind Instruments Synthesis Toolbox (Rocamora, López, & Jure, 2009) and Praat (Boersma & Weenink, 2001). The Wind Instruments Synthesis Toolbox was used to create two 500ms musical instrument notes, one synthesized from a trumpet template and one synthesized from a trombone template. Both notes were synthesized with identical fundamental frequencies and identical intensity properties; as such, the only thing distinguishing the two notes was their timbre.

Next, the notes were spectrally rotated, a type of acoustic manipulation that redistributes information across frequencies in an acoustic signal. Within a speech context, spectral rotation is often seen in neuroimaging studies, where it is used as a way to create a signal with much of the acoustic richness of speech but without the phonetic, syntactic, or semantic properties of speech itself (Peelle, Gross, & Davis, 2013; S. K. Scott et al., 2000). Spectral rotation was used in an analogous sense here to construct synthetic "musical instruments" that have much of the rich acoustic signature of brass instruments but without a true connection to the instruments. This renders them analogous to the German fricatives in the phonetic category learning experiments: acoustically complex and clearly "instrumental" but unfamiliar. The trumpet and trombone sounds were spectrally rotated using two channels (split at 8000 Hz) to create two endpoints for my musical instrument continuum, which were labeled the "pettrum" and the "bonetrom", respectively. These tokens were peak scaled to ensure their intensities matched. Next, Praat was used to linearly combine the two endpoints in order to make a 10-step continuum. Just as with the speech stimuli used in the prior experiment, this was accomplished through use of spectral blending: each point along the continuum represented a linear combination of the two endpoint signals. The pettrum end was arbitrarily labeled Step 1, while the bonetrom end was arbitrarily labeled Step 10. Step 8, then, to pick one arbitrarily, was primarily comprised of a bonetrom signal, but with some pettrum properties intermixed.

2.3.2 Experiment 2: Discriminability

To make inferences from a direct comparison of the instrument and fricative materials, it was necessary to get a sense of the properties of the materials used. After all, any differences that would be found between the acquisition of phonetic and instrument categories could either be the result of differences in the processing of items inside and outside of language (the main object of study in the present experiments) or simply due to differences in the acoustic properties of the stimuli. Secondarily, I was interested in whether participants perceive both continua in a unidimensional fashion, as largely assumed in previous treatments of the fricative stimuli, or if they perceived them using multiple dimensions. I used Amazon's Mechanical Turk service to recruit participants for a simple study where participants made stimulus similarity judgments for stimulus pairs spanning the continua, described below.

2.3.2.1 Participants

27 participants were recruited from Amazon's Mechanical Turk crowdsourcing database. One participant was thrown out for experience with German, leaving 26 native English speakers (7 female, 19 male). Although the participants generally skewed older than a typical undergraduate population (*M* = 34.7, Range = 25-47), none were old enough that high-frequency hearing loss would be expected. Although headphone use was requested for the task, 3 participants reported using external or built-in speakers. Despite uncertainty about the precise qualities of the sound equipment that the participants used, previous studies using Mechanical Turk (Buxó-Lugo & Watson, 2016; Heffner, Newman, & Idsardi, 2017; Slote & Strand, 2016) have generally found Mechanical Turk to be an appropriate venue to run experiments related to phonetics.

2.3.2.2 Materials

Participants heard two blocks of trials: one using the non-speech stimuli described above, and another using the fricative stimuli in previous category learning experiments. The order of each block was counterbalanced across participants.

2.3.2.3 Procedure

Participants heard two paired stimuli from one of the continua, back-to-back. With ten possible stimuli as both the first and second item, there were therefore 100 possible pairs of stimuli per continuum. Participants heard all 100 pairs exactly once, and were then asked to rank how similar the items within the pair were on a scale from 1 to 9.

2.3.2.4 Analysis

The similarity judgments for each participant were converted into difference scores, ranging from 0 (not different) to 8 (most different). These difference scores were used to create a 10 × 10 symmetric data matrix for each participant, with each row and each column being a step within the continuum. These symmetric data matrices were analyzed using the IDIOSCAL (Individual Differences in Orientation Scaling) functionality of the “smacof” package within R (Mair, De Leeuw, Borg, & Groenen, 2016). IDIOSCAL is a generalization of Individual Differences Scaling, INDSCAL (Carroll & Chang, 1970), which has been used extensively in the category learning literature; for example, in determining naïve listeners’ parcellation of Mandarin tone categories (Chandrasekaran, Sampath, & Wong, 2010) or to examine the effects of training on categorical perception (Livingston, Andrews, & Harnad, 1998).

In INDSCAL and IDIOSCAL, dimensionality analysis requires multiple possible dimensionalities, *n*. For each dimensionality, an *n* by 10 matrix is generated, showing the coordinates of each stimulus step in an *n*-dimensional space. Traditionally, the approach to determine the best number of effective dimensions is to calculate badness-of-fit measures for each *n* and to look for an "elbow", a point at which additional possible dimensions do not lead to appreciable drops in badness ratings.

2.3.2.5 Results

Participants by and large perceived both continua as unidimensional. Figure 9 shows a scree plot with badness-of-fit values across different possible dimensionalities. Higher stress values indicate larger badness-of-fit. The lines do not show a clear "elbow"; badness-of-fit decreases gradually across the possible dimensionalities for both continua. Although the largest numeric difference across dimensionalities occurs between one and two dimensions (0.049 for the fricatives, 0.076 for the instruments), that difference is not particularly large nor much bigger than the next largest difference, between two and three dimensions (0.030 for the fricatives, 0.033 for the instruments). I find no evidence to reject the unidimensional interpretation of the continuum.

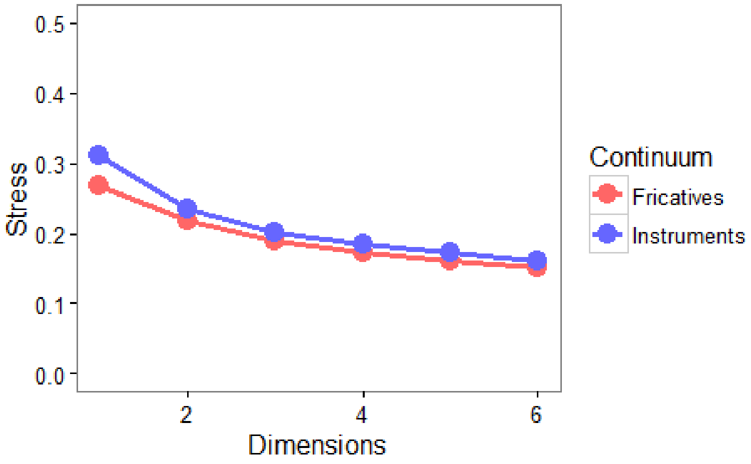


Figure 9. Badness-of-fit values in Experiment 2

To the extent that the stimulus similarity ratings did not conform to a unidimensional distribution, in fact, participants generally found the endpoints of the continuum to be more similar to each other than would be expected given a uniform progression from most to least similar items. This was true to reasonably similar extents for the fricatives and the instruments sounds. This can be seen in Figure 10, below, which shows the two-dimensional IDIOSCAL solution.

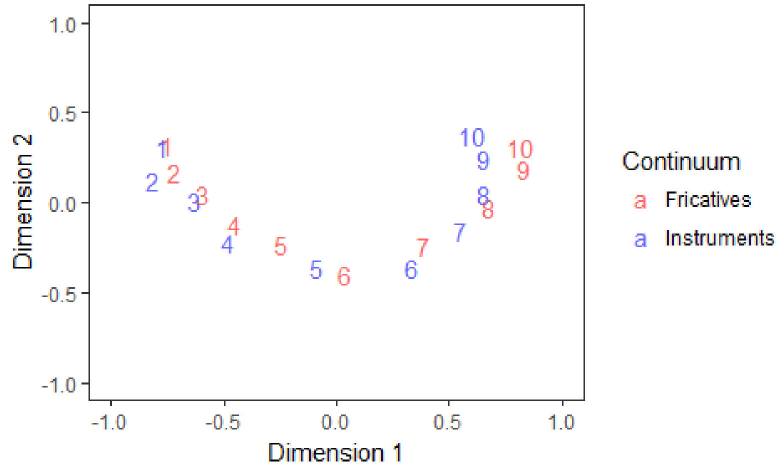


Figure 10. Two-dimensional IDIOSCAL solution, Experiment 2

The dimensions revealed in Figure 10 roughly correspond to the position in the stimulus along the continuum (Dimension 1) and to whether the stimuli are extreme members of the continuum or fall somewhere in the middle (Dimension 2). To put it another way, Dimension 1 showed which category that naïve English speaking participants sorted the items into, while Dimension 2 showed the level of certainty that the participants had in that label (with higher values indicating increasing certainty). As such, the “extremely palatal” items (steps 1-3) and the “extremely velar” items (steps 8-10) are less distant from each other than one would expect based on stimulus step alone, as listeners were very certain about the categorization of both endpoints. In general, the items are classified similarly across the two conditions, with roughly equal distances from step to step across the two continua. This suggests that comparing the two conditions is appropriate (although see the Discussion for additional speculation about this).

2.3.3 Experiment 3: Learning Musical Instrument Categories

Given that the continua appeared to be comparable, the next step was to actually test the category learning for the instrumental materials. To do this, I used a paradigm essentially identical to that used for the fricative categories, with the only difference being in the materials used.

2.3.3.1 Participants

63 participants were enrolled in the experiment. Of those, 8 participants were excluded from further analysis: 1 because of a missing demographics survey, 3 due to technical errors, and 4 due to a failure to follow directions (as indicated either by an unusual response strategy[2] or 10 or more trials without a timely response). That left 55 participants with analyzable data (27 female, 28 male). All participants were at least 18 years of age (*M* = 20.5, Range = 18-29) and had no history of hearing impairments. Participants were recruited from the University of Maryland, College Park community for either course credit or a $10/hour compensation.

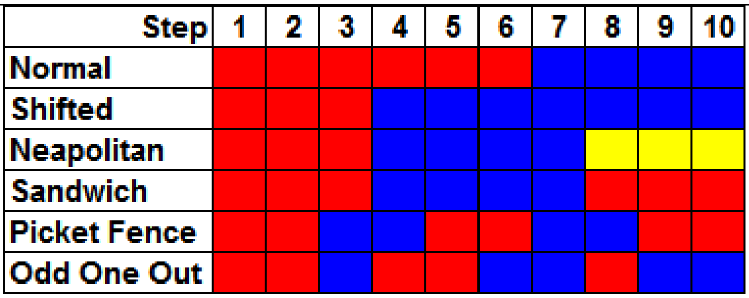
2.3.3.2 Procedure

The procedure used in this experiment was identical to the one used in Experiment 1. First, participants heard one of the sounds from the instrument continuum.

[2] This participant pressed every single key simultaneously on every single trial until corrected.

Next, they pressed a button corresponding to one of three colored squares along the continuum: blue (“J”), yellow (“K”), or red (“L”). Finally, they received feedback on their selection, a green check if their response is correct, and a yellow “X” if their response is incorrect. Participants were told to let the feedback they get on a trial-by-trial basis guide their responses. The only difference in the procedure was that the instructions changed to tell participants they were learning to pair generic “sounds” with colored squares rather than “speech sounds” in particular.

What differed from participant to participant is which of six conditions that people are assigned to, and, therefore, which responses are considered “correct” or “incorrect”. The conditions used, which were identical to those used in Experiment 1, are illustrated in Figure 11. Each colored square shows the correct response for each step in each condition. For example, the correct response for Step 8 in the Normal, Shifted, and Picket Fence conditions is to press the button corresponding to “Blue”, while the correct response in the Sandwich and Odd One Out conditions is “Red” and the correct response in the Neapolitan condition is “Yellow”. Again, all three responses were available to the participants across the conditions.



Step	1	2	3	4	5	6	7	8	9	10
Normal	🟥	🟥	🟥	🟥	🟥	🟥	🟦	🟦	🟦	🟦
Shifted	🟥	🟥	🟥	🟦	🟦	🟦	🟦	🟦	🟦	🟦
Neapolitan	🟥	🟥	🟥	🟦	🟦	🟦	🟦	🟨	🟨	🟨
Sandwich	🟥	🟥	🟥	🟦	🟦	🟦	🟦	🟥	🟥	🟥
Picket Fence	🟥	🟥	🟦	🟦	🟥	🟥	🟦	🟦	🟥	🟥
Odd One Out	🟥	🟥	🟦	🟥	🟥	🟦	🟦	🟥	🟦	🟦

Figure 11. Correct responses in Experiment 3

The training was brought to a halt under two conditions. First, participants were judged to have learned the pairings of sounds to squares if they got the most recent appearance of 9 out of the 10 stimulus steps correct. If this was achieved, the experiment ended immediately, and the number of trials that had elapsed was taken as a guide for how difficult the condition was to learn, with a greater number of trials indicating a more challenging condition. Alternatively, the experiment ended after 450 trials, no matter what the participants' responses were, as a way to keep the experiment from continuing endlessly.

2.3.3.3 Analysis

Two analyses were used to assess participants' learning of the categories in question. As before, one of those analyses comes from the number of trials necessary to learn the pairings of sounds to colored squares, which served as a time-to-criterion measure. As in the phonetic category learning experiments, participants were divided into fast learners (less than 225 trials to learn), slow learners (more than 225 trials), and non-learners (no learning within 450 trials), and the number of participants in each group was compared across conditions.

The second analysis involved model comparison using generalized linear mixed models (Bates et al., 2016). This analysis included the data from the participants run using the fricative continuum, and started with a model that had a mixture of fixed and random effects, all with the goal of predicting trial-by-trial variation in accuracy (coded in a binary fashion, with 1 as a correct answer and 0 as an incorrect one). The fixed effects included in the initial model were the condition (Normal, Shifted, Neapolitan, Sandwich, Picket Fence, and Odd One Out), continuum (Fricative and Instrumental), and the continuous factor of trial number (rescaled and zero-centered to range from -0.5 to 0.5; trial number 225 was the 0 point). As such, fast-learning participants generally only had values of trial number that were below zero; only slow learners had trial number values that were greater than 0. The random effects were intercepts by participant and by step and random slopes for trial number by participant (to model variation in learning behavior) and for condition by stimulus step (to model differences between items in learnability across the conditions). This model was then compared to models with simpler fixed effects structures to determine significant fixed effects. The best-fitting model was then used to compute post hoc measures of differences between conditions and between the fricative and instrumental stimuli using the lsmeans package (Lenth, 2016).

2.3.3.4 Results

The results for this experiment are given below in Figure 12. The number of trials needed to learn each condition (horizontal axis) in the musical instrument category learning experiment, graphed by condition and continuum (marked at left). The data from the English-speaking participants in Experiment 1 is presented again here for comparison. Each participant is shown as a grey circle, with red crosses showing the median time to learn for participants who successfully learned that particular condition. Participants clustered along the vertical line at the right did not successfully learn the pairings within 450 trials. Vertical jitter away from the vertical lines is meant to make individual participants clear.

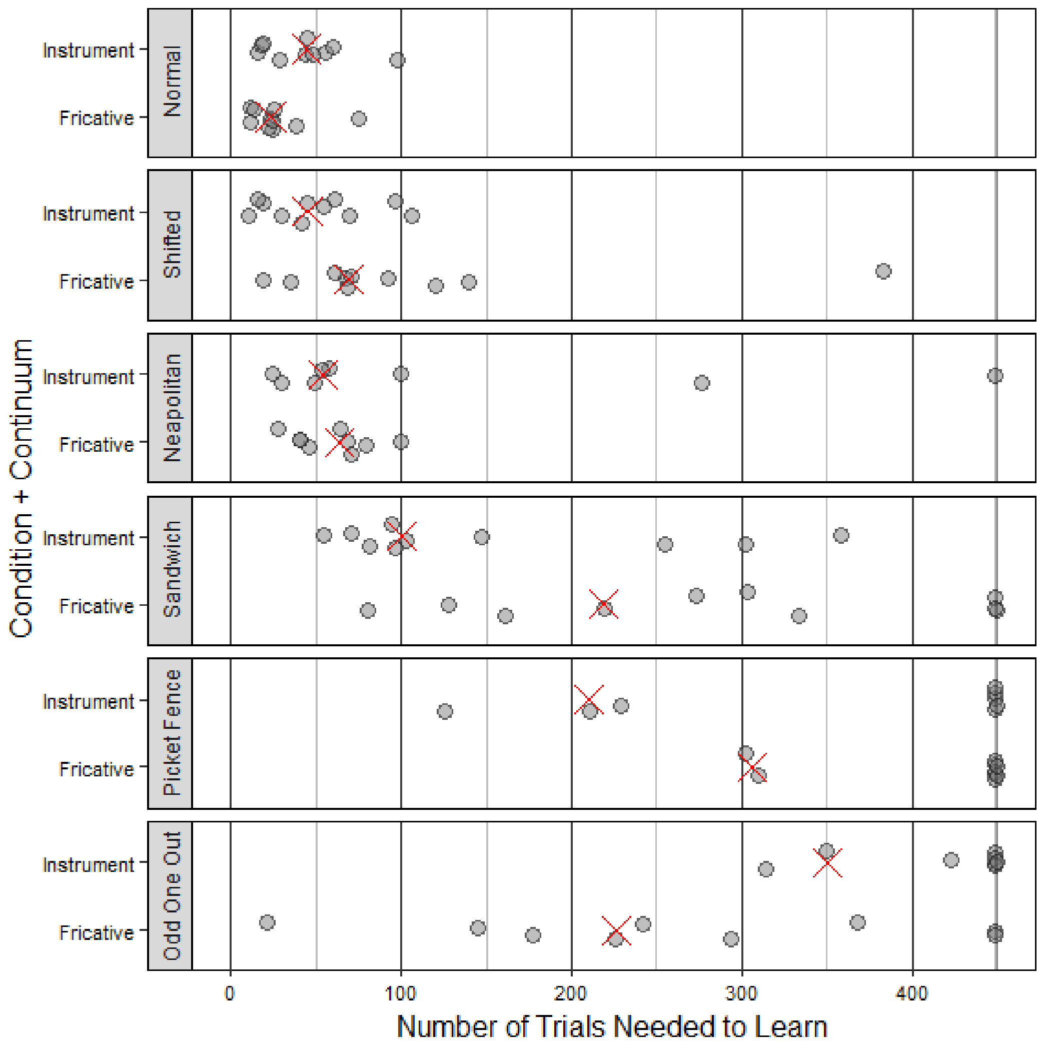


Figure 12. Experiment 3 results by participant

At first glance, the results for this experiment resemble those found for phonetic categories: the Normal and Shifted conditions are easy to master, while the Picket Fence and Odd One Out categories are difficult. However, a more detailed examination of the results yields a surprising finding: the distinction between the Neapolitan and Sandwich conditions shrank, quite strongly, for the non-speech categories. This can perhaps be seen more clearly in Table 3, below, which compares English speakers learning fricatives to English speakers learning instruments in their median times to learn for each condition.

The results for the fricative continuum are for native English speakers only. Again, the median times to learn are presented solely for the sake of description.

	MTTL	
	Fricatives	**Instruments**
Normal	23.5	44.5
Shifted	69.5	45
Neapolitan	64	56
Sandwich	219	100
Picket Fence	306	210
Odd One Out	226	350

Table 3. Experiment 3 result summary tables

This change appeared to be largely dependent on the Sandwich condition. That condition went from being learned most of the time for the fricative continuum (70%) to being learned all of the time for the instrument continuum (100%). It also went from being learned quite slowly for the fricatives (a median of 219 trials to learn) to much more quickly for the instrument continuum (a median of 100 trials to learn). The Sandwich condition now more closely resembles the Neapolitan condition; it is “easy” where it once was difficult.

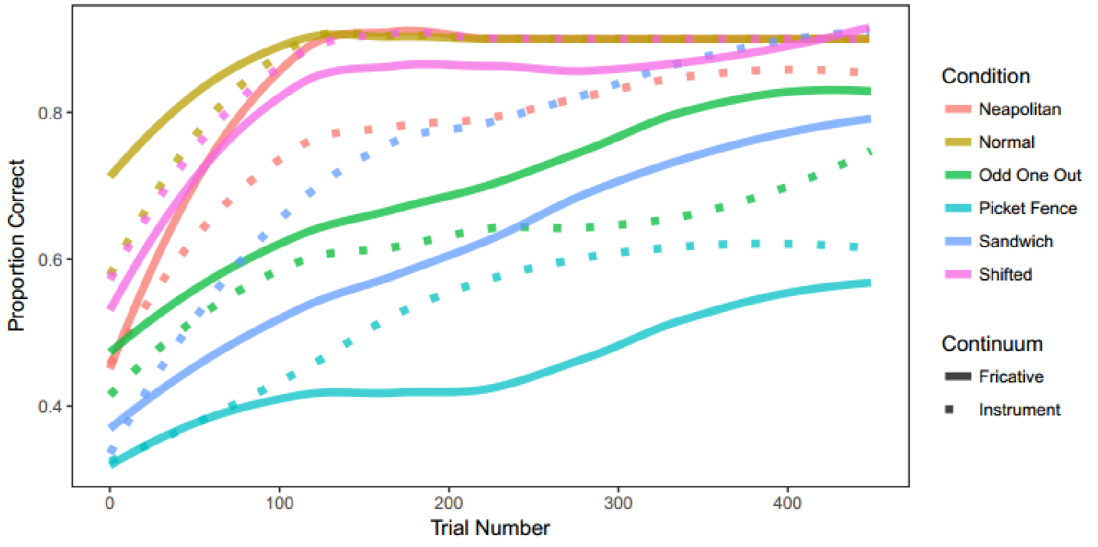


Figure 13. Experiment 3 results, Loess-smoothed

These effects are echoed by the mixed model analysis of the results. Loess-smoothed learning trajectories, with the proportion correct for participants who successfully learned the pairings of sounds to squares padded with a value of 0.9 through trial number 450, are depicted in Figure 13. The best fitting model that was one that included all of the fixed effects in general, including both two-way and three-way interactions, which is described in detail in Table 4 (using the fricatives as the reference level for continuum and the Neapolitan condition as the reference level for the condition). Comparing the best-fitting model to one without the fixed effect of condition (and its interactions with other fixed factors) yielded a significant decrease in model fit without the condition, $\chi^2(20) = 97.9$, $p < .001$, indicating that some conditions had a higher baseline acceptance than others. Removing trial number also significantly decreased model fit, $\chi^2(12) = 129$, $p < .001$; participants did learn over time. And taking away the effects of continuum also made a difference in learning, $\chi^2(12) = 29.9$, $p = .003$; participants differed in their success depending on the continuum they were trying to learn.

All of the interactions in the best-fitting model also contributed a significant amount to the model fit. Removing the interaction between the continuum and the trial number (and the three-way interaction) led to a significant decrease in model fit, $\chi^2(6) = 24.7$, $p < .001$, indicating that there was a difference in the pace of learning according to the continuum. Removing the interaction between the continuum and the condition (and the three-way interaction) also led to a decrease in model fit, $\chi^2(10) = 28.7$, $p = .001$, indicating that the different conditions had different baseline learning between each continuum. And, finally, removing only the three-way interaction between continuum, condition, and trial number also led to a significant drop in model fit, $\chi^2(5) = 24.5$, $p < .001$, meaning that the differences in the rate of learning across conditions were contingent on the continuum. That is, which conditions were easy to learn depended on the continuum of sounds being learned.

Fixed Effect	***β***	**SE**	***z***	***p***
Intercept	5.34	1.21	4.40	<.001
Condition: Normal	0.734	2.32	0.32	.75
Condition: Shifted	-3.51	1.31	-2.67	.008
Condition: Sandwich	-5.08	1.22	-4.16	<.001
Condition: Picket Fence	-5.62	1.23	-4.56	<.001
Condition: Odd One Out	-4.88	1.32	-3.70	<.001
Trial Number	11.6	2.73	4.25	<.001
Continuum: Instrumental	-4.62	1.28	-3.60	<.001
TN × Cnd: N	-0.22	5.01	-0.044	.96
TN × Cnd: Sh	-8.73	2.92	-2.99	.003
TN × Cnd: Sa	-10.3	2.76	-3.72	<.001
TN × Cnd: PF	-11.0	2.77	-3.96	<.001
TN × Cnd: OOO	-10.5	2.77	-3.79	<.001
Cont: I × Cnd: N	5.66	2.50	2.27	.02
Cont: I × Cnd: Sh	6.51	1.64	3.97	<.001
Cont: I × Cnd: Sa	5.03	1.33	3.78	<.001
Cont: I × Cnd: PF	4.84	1.33	3.64	<.001
Cont: I × Cnd: OOO	4.60	1.32	3.48	<.001
TN × Cont: I	-9.99	2.91	-3.43	<.001
Cont: I × Cnd: N × TN	13.3	5.42	2.46	.01
Cont: I × Cnd: Sh × TN	14.4	3.72	3.89	<.001
Cont: I × Cnd: Sa × TN	10.8	3.02	3.60	<.001
Cont: I × Cnd: PF × TN	10.5	2.99	3.51	<.001
Cont: I × Cnd: OOO × TN	9.82	2.98	3.30	<.001

Table 4. The best-fitting model for Experiment 3

The lsmeans package was used to inspect differences between the conditions for the instrument continuum and differences between the fricative and instrument continua. In particular, three targeted tests were used to match those performed in the fricative continuum: a comparison between the Normal and Shifted conditions, a comparison between the Normal and Odd One Out conditions, and one between the Neapolitan and Sandwich conditions. Additionally, post hoc tests were performed to compare learning in the instrument continuum with learning in the fricative continuum. The multivariate *t* adjustment was applied to ensure that multiple comparisons reflected a truer measure of significance.

2.3.3.4.1 Normal vs. Shifted

There was no a priori reason to expect the Normal and Shifted conditions to differ in the rate of learning for the musical instrument continuum. And, indeed, that was the case; the lsmeans package indicated that the Normal and Shifted conditions did not differ in the rate of learning, $z = 2.16$, $p = .23$.

2.3.3.4.2 Normal vs. Odd-One-Out

In contrast to the fricative continuum for English speakers, the Normal vs. Odd One Out contrast for the instrumental continuum was significant. Indeed, the Odd One Out condition was much harder than the Normal one, $z = 4.77$, $p < .001$. Learning was a good deal slower for participants in the Odd One Out condition.

2.3.3.4.3 Neapolitan vs. Sandwich

Unlike for the fricatives, however, there was no difference between the Sandwich and Neapolitan conditions in learning for the instrumental continuum, $z = 0.313$, $p = 1$. The Sandwich condition was no slower (or, at least, not significantly slower) than the Neapolitan condition in the rate of learning for the fricative continuum. Additionally, the participants who took longer in the Sandwich condition for the instruments were not performing similarly to the participants in the Sandwich condition for the fricatives. Figure 14 shows participants' performance during the last 25% of trials in the Sandwich condition. Each row corresponds to the behavior of a single participant, with 10 columns that correspond to the responses participants gave. Each cell is coded in line with the proportion of responses from each category. Cells that are entirely blue, red, and yellow indicate that participants responded 100% with that color category for that step within the last 10% of trials administered, while cells with intermediate colors represent combinations of responses. For example, purple cells represent some red and some blue responses for those steps towards the end of the experiment. The white cell for participant ME22 indicates no responses for that stimulus step in the last 25% of trials. As can be seen, participants were only very rarely using the yellow button to respond, unlike for the fricatives, and most of the participants seemed to have little trouble positing a red category that was on both ends of the continuum.

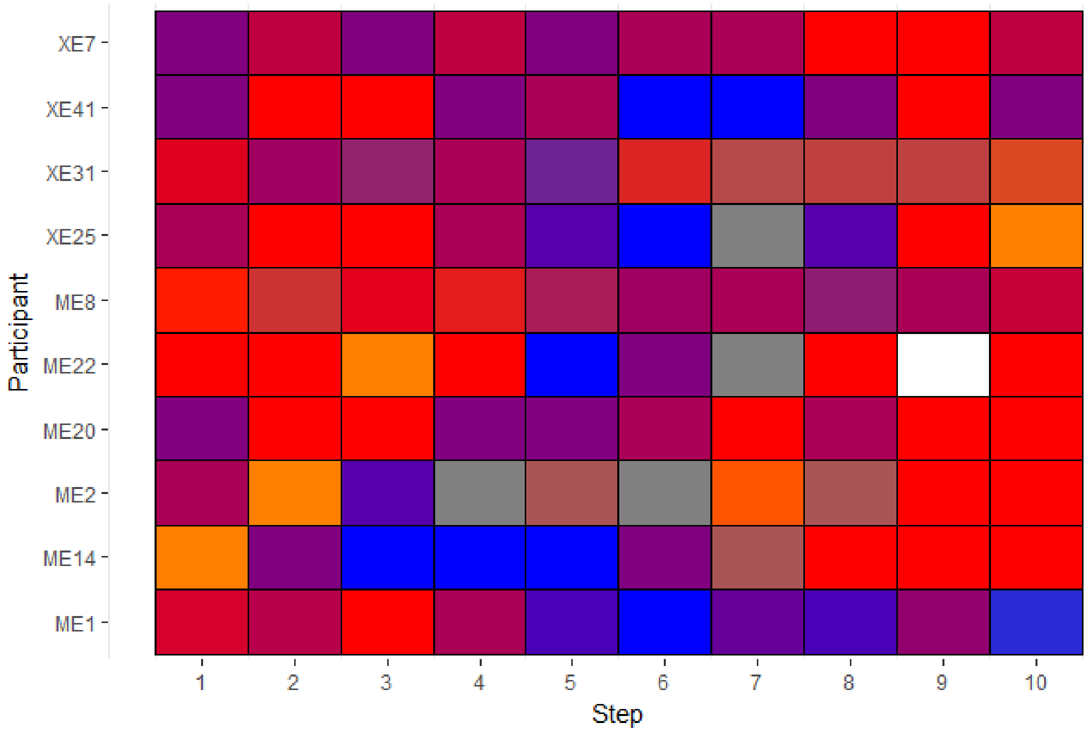


Figure 14. Experiment 3 Sandwich final section results

2.3.3.4.4 Fricatives vs. Instruments

A cursory comparison of Figure 6 and Figure 12 suggests that there are no differences between the fricative and instrument continua in the rate of learning for the Normal and Shifted conditions (which are easy for both continua), nor for the Odd One Out and Picket Fence conditions (which are difficult for both continua). And, indeed, post hoc comparison of the fricative and instrumental continua shows that none of these comparisons reach significance: comparisons of the Normal, *z* = -0.370, *p* = 1, Shifted, *z* = -1.79, *p* = .46, Odd One Out, *z* = -0.010, *p* = 1, and Picket Fence, *z* = -0.625, *p* = 1, conditions show no differences in the time to learn according to the continuum.

The results for the Sandwich and Neapolitan conditions are somewhat more surprising. Although it may seem like the distinction between the instrumental and fricative stimuli is driven by the numerical difference in the Sandwich condition, such that the Sandwich condition is easier to learn in the instrumental continuum than the fricative continuum, this difference is not significant, *z* = -1.12, *p* = .91. The main difference is in fact in the Neapolitan condition, where learning in the instrument continuum is slower than in the fricative continuum, *z* = 3.63, *p* = .003. This is perhaps caused in part by the long tail of the Neapolitan condition for the instruments. In contrast to the uniformly speedy learning of the Neapolitan condition for the fricatives, there were a few participants in the instrumental continuum who took more than 100 trails to learn to pair the instruments with the colored squares.

2.3.4 Discussion

To summarize, then, it appears to be the case that disjunctive categories were no harder to learn than non-disjunctive categories for the instrumental continuum. The post hoc comparisons indicated that this failure to find a difference was largely the result of the Neapolitan condition challenging the participants more for the instruments than for the fricatives.

These findings do not definitively rule out the existence of an anti-disjunctivity bias for the instrumental continuum. There was a trend in the same direction that was observed in Experiment 1: the Sandwich condition was slightly harder to learn than the Neapolitan one. Although the difference between the Neapolitan and Sandwich conditions was not significant in Experiment 3, the small sample size of the groups in the present experiment suggests that power may also have been an issue; perhaps the analyses did not uncover an effect simply because the participant numbers were too small. Of course, the fact that such an effect was uncovered using an identical sample size in Experiment 1 suggests that, at the very least, such an effect would be relatively weak for non-speech sounds when compared to speech sounds. Yet even a weak effect would suggest that the bias against disjunctive categories observed in Experiment 1 is not specific to speech. The implications of this work would then shift to why the bias against disjunctive categories was *larger* in speech than non-speech; which, though an interesting finding, would tamper down on some of the stronger theoretical issues of the present project. Although exploring this possibility in greater detail is outside the scope of this paper, planned follow-up experiments involve running participants online (to increase the speed and ability of data collection) and requiring all participants to complete the same, large number of trials (to ensure that a large amount of data is collected for each participant), which should help address the sample size concerns in the present experiment.

An alternative explanation for the failure to find an effect would implicate a performance ceiling. As such, the failure to find an effect would stem from the participants in the Sandwich and Neapolitan conditions both reaching this ceiling. However, this seems unlikely, given that the performance in the Odd One Out and Picket Fence conditions was much poorer than either the Sandwich or the Neapolitan conditions, and that the continua were quite closely matched to each other in terms of their perceptual properties.

All told, however, I believe that the experiments in the present dissertation suggest that learners may be subject to different biases when learning speech sound categories from those that may exist for non-speech analogues. The dataset is discussed below in terms of the theories of category learning sketched in the review chapter and the theories of domain-specificity in speech perception discussed in the introduction.

2.3.4.1 Relevance to Theories of Category Learning

2.3.4.1.1 Exemplar-Only

The results of the present experiment are easier to integrate with exemplar-only theories of category learning than the results using the fricatives. As expected, the Normal and Shifted conditions were easy for participants to learn; no participants took longer than 125 trials to achieve the learning criterion in those trials. Conversely, the Odd One Out and Picket Fence conditions were very challenging, with few participants successfully acquiring the categories in either condition, let alone acquiring them quickly. And, finally, the Neapolitan and Sandwich conditions were equivalently difficult (i.e., not very challenging), in contrast to the results obtained for the fricative continuum, where there was a strong split between the two conditions. The labels of the continuum endpoints for the instruments did not have a significant influence on the rate of learning. This is exactly in line with the predictions of exemplar-based category learning.

Reconciling both the fricatives and the instruments, however, is more challenging. Under an exemplar-only system where identical learning systems are used inside and outside of language, it can be hard to explain why learners should be hobbled by disjunctive categories in one domain that they have no problem learning in another, very similar one. Indeed, performing both experiments, and finding this dissociation in one domain and not the other, rules out one tempting possible explanation: the idea that the Sandwich condition involves participants learning to ignore the yellow response key. Participants in the Sandwich condition for both the fricative and instrumental items had to ignore the yellow response key; but only for the fricatives did this trip them up.

Using exemplar-only mechanisms to explain these results is not impossible, however. One possible explanation relates to the previous experience that listeners had with the sounds in question. The German-speaking participants in the previous experiments had extensive experience with the fricatives, while the English speakers often had at least sporadic experience with the sounds in question. This was not true for the instrumental sounds used; as the bonetrom and pettrum sounds were created for this experiment, none of the participants came in with any exposure whatsoever. This explanation, though, seems quite unlikely. Despite the fact that both German and English speakers may have had *some* experience with the fricatives in general, it is also indisputable that the German speakers had orders of magnitude more experience with them. Yet both participant groups showed the bias against disjunctivity in learning despite those differences. Furthermore, it is the English speakers, not the German speakers, who had the largest difference between the Sandwich and Neapolitan conditions in the time it took them to reach the learning criterion; this would imply that having either too much experience (as with German speakers learning German fricatives) or too little experience (as with English speakers learning instrument sounds) would shrink the difference between disjunctive categories and their non-disjunctive equivalents. This is a possible outcome, but it does not seem likely.

Another possible exemplar-based explanation for the differences between the disjunctive categories and the non-disjunctive categories relates to differences in the dimensionality of the stimuli. To see why, consider the role of stimulus dimensions in learning under exemplar-based models. Under most exemplar-only models of category learning, learners are not just acquiring the categories being introduced to them as a part of the experiment; they are also learning what weight to place on the different physical dimensions that distinguish the categories. This can be seen in models such as ALCOVE, where flexible attentional weights drive participants' reliance on different dimensions (Kruschke, 1992). These attentional weights could, theoretically, be preset, or depend on the physical properties of the dimensions being sampled. The attentional weights change the perceptual distance between two items, which, in turn, affects the rate of learning for those items. Although this has little influence on unidimensional category learning, in situations for which there is evidence for multiple dimensions, differences between the attentional weights across those situations could lead to differences in the pace of learning.

For the category learning experiments discussed here, then, such an argument depends on several premises. One must discard the conclusion reached as a result of Experiment 2 that both stimulus continua are perceived in a unidimensional fashion, at least for the fricative items. Furthermore, these multiple dimensions must be dissociable enough that participants can allocate different levels of attention to the ones that are relevant to the categories at hand. One of these dimensions should be allocated the lion's share of attention at the beginning of the experiment, but that dimension should make it hard to distinguish between items, such that learning the disjunctive tokens would be impossible using that dimension only. This would keep performance at the beginning of the experiment quite low, as it was with the Sandwich condition in both continua. Another dimension, meanwhile, must be capable of distinguishing the category endpoints from the middle of the continuum, and that dimension should be allocated less attention at the beginning of the experiment than others. Over the course of the experiment, a shift in the allocation of attention to this second dimension would lead performance to improve, because the second dimension could lead to a correct demarcation of the continuum into two categories.

If the dimensions shown in Figure 10 are dissociable, they provide an example of the type of setup that would be necessary for at least the instrumental items. Recall that Figure 10 showed the results of a two-dimensional IDIOSCAL solution to untrained participants' perception of the fricatives and instrumental sounds. For both continua, Dimension 1 appears to relate to the physical properties of the sound stimuli, whether that is cues to the place of the fricatives or cues to the instrument identity for the instruments. Dimension 2, meanwhile, seems to relate to extremeness in both continua; stimuli with positive values on Dimension 2 are at the ends of each continuum, while stimuli with negative values are towards the middle.

With these dimensions, it is possible to construct a situation in which an exemplar-only learning system takes longer to acquire the Sandwich condition than the Neapolitan one. If the learner starts with a strong bias towards attending to Dimension 1, and Dimension 1 is insufficient to learn the categories in the experiment, learners should find the beginning of the experiment hard. Over time, then, they should shift their attention to Dimension 2, which provides a very simple way to distinguish red items (at the endpoints) from blue items (in the middle). This will take time, but ultimately lead to learning. The idea is that this strong bias towards attending to Dimension 1 could be present for fricatives but absent for the instruments.

Some aspects of this idea are tempting. It does not seem odd to suppose that the perceptual dimensions that are used to distinguish the fricatives from one another are different from those used to distinguish the musical instruments from each other. Further, it is reasonable to suppose that listeners may be biased towards weighting some of those dimensions stronger for the fricatives than for the instruments. However, the full argument is not very tenable. As discussed in greater deal above, Experiment 2 provides weak evidence, if any, that the category learning problem here is two dimensional. More worryingly for exemplar-only theories, it is not clear why there would be a difference between the Neapolitan and Sandwich conditions in attentional weighting at the onset of learning that would drive participants in the Sandwich condition to reallocate attention while Neapolitan participants do not. That is, if using only Dimension 1 to categorize items in the Sandwich condition is prevented by the inter-item confusability in that condition, there is no reason why using only Dimension 1 to categorize items in the Neapolitan condition would be any easier. They are the same items with the same boundaries between them.

Testing this idea would require additional knowledge about the dimensionality of the stimuli used here. One could imagine, for example, performing a similarity judgment task similar to the one used in Experiment 2 both before and after category learning. If learning is driven by changes in dimensionality, participants' similarity judgments should reflect the attentional shifts required to learn the disjunctive categories. To sum up, taking an exemplar-only approach to explain a bias against disjunctive categories is challenging. The most plausible recourse is to re-allocate attention from salient dimensions that prevent successful learning to less salient dimensions that allow for learning; however, such a pathway seems unlikely.

2.3.4.1.2   Dual-System

For dual-system models, meanwhile, the differences in learning between the continua are easier to accommodate. It might be the case, for example, that the rule-based category learning system has different properties for phonetic and non-phonetic stimuli such that it is constrained to positing non-disjunctive categories for speech in a way that it is not required to do for non-speech stimuli. In other words, the "over-hypotheses" of phonetic learning would be different from the "over-hypotheses" of non-phonetic learning (Kemp et al., 2007). Learners may be biased in this way because of the relative frequency of disjunctive categories in language versus disjunctive categories in music. Disjunctive categories are rarely present in language, barring occasional counterexamples, as with allophones of /t/ and inter-individual variation in the pronunciation of /ɹ/ and /ʃ/. They are much more frequent, though, in music, as categories such as "C" refer to a variety of widely scattered pitches, one per octave. It would also be reasonable to propose that the similarity-based system is slower to activate for the fricative items than the instrument items, which would in turn lead to slower learning for the disjunctive categories.

Perhaps one of the biggest weaknesses of dual-system models is the sheer number of ways in which learning might differ between the fricatives and instruments. Unlike in exemplar-only models, where essentially only two factors—the dimensionality of the stimuli being used, and the amount of noise in the perceptual system that can distinguish between items—can affect the acquisition of categories, the multiple systems that characterize dual-system models can vary in many ways within each system. Honing in on any one explanation is challenging. Distinguishing between, say, changes within a single system versus changes in the handoff of control from one system to another require follow-up studies that selectively disrupt the activity of each purported system (Maddox & Ashby, 2004). Thus, while this pattern of results could be explained by a dual-system theory of category learning, the system would need to be of a type that discourages the acquisition of disjunctive categories in the rule-based system, as in RULEX (Nosofsky et al., 1994); and, curiously, that bias would need to manifest only for certain types of category learning contexts.

2.3.4.2 Relevance to Domain-Specificity

This study also has direct relevance to the idea of domain-specificity in speech perception. The acquisition of phonetic categories seemed to be constrained in a way that the acquisition of non-linguistic categories was not. This is relatively easy to accommodate under domain-specific theories of speech perception, where the process of acquiring new phonetic categories can have fundamentally different properties from the process of acquiring other acoustic categories. However, without recourse to previous experiences with items similar to each continuum, it is more challenging to integrate with domain-general theories of phonetic perception, which would predict just the opposite. Examples of each approach are sketched below.

2.3.4.2.1 Domain-Specific Theories

Motor theories of speech perception have little difficulty accommodating differences in learning between phonetic perception and other types of auditory perception. Under motor theories, phonetic signals, once their “phonetic” nature is detected, are processed in a special way through the reconstruction of underlying motor gestures. Other acoustic signals are processed separately. Given that phonetic perception is domain-specific under motor theories, it is entirely possible to accommodate different learning processes within a domain from those outside of it.

Indeed, a bias against disjunctive categories may fall out quite neatly from the predictions of motor theories. If the underlying perceptual dimensions at work for speech sounds are motor, not perceptual, a dispreference for disjunctive categories makes some sense. Motor gestures are often quantal (a lip motion or tongue gesture is made or not made), which should work against positing theories that involve two separate, disconnected motor gestures being a part of a single category. However, there are some phonetic categories that could be reasonably proposed to be disjunctive in terms of motor actions, even just in American English. [ɹ] can be produced using a huge variety of motor gestures, involving both the tongue tip and the back of the tongue (Guenther et al., 1999; Westbury, Hashi, & Lindstrom, 1998). [s] and [ʃ] are also produced with a variety of articulatory configurations, with individual participants showing only limited overlap between the categories in production but with between-participant overlap being substantial (R. S. Newman, Clouse, & Burnham, 2001). As such, it is not clear whether motor theories speak to the idea of disjunctive categories in speech, though they would of course easily accommodate differences between the speech and non-speech categories.

Acoustic theories of speech processing that incorporate speech-specificity should also have few challenges accommodating this distinction. As with motor theories, these theories propose that speech-like acoustic information is sent to a specialized acoustic module in order to enter into later linguistic computations (Poeppel et al., 2008). If speech processing has a certain set of special, enumerated features that make it different from other types of perceptual processes, it is not hard to include a dispreference for disjunctive categories into that list. This becomes particularly relevant under theories that postulate abstract distinctive features. Consider the distinction between palatal and velar fricatives used in the phonetic learning experiments in the previous chapter. Under those theories, palatal and velar fricatives are distinguished from one another by the presence of different place features, such as [+/-high] and [+/-low]. One could split the fricatives from one another into separate categories by emphasizing the place features that divide them, or combine them into a single category by deemphasizing those same features. However, it is unclear how to posit a category that includes both the palatal and velar endpoints without including the items in between. For example, consider the table of place distinctions and features in Table 5 below, where Distinctive Feature 1 (DF1) and Distinctive Feature 2 (DF2) characterize the difference between the palatal and velar fricatives, while intermediate items are characterized by the presence of positive values for both features.

	Palatal	Intermediate	Velar
DF1	+	+	-
DF2	-	+	+

Table 5. A toy feature list

In a distinctive features account of category learning, it is straightforward to learn three different categories along the continuum; the red category is defined as any items with [+DF1, -DF2], the yellow category as [-DF1, +DF2], and the blue category as [+DF1, +DF2]. But it is not possible to easily characterize the palatal and velar fricatives in a way that is distinguishable from the intermediate fricatives; the palatal and velar fricatives do not have any featural specifications that are unshared with the items in between them. This is a relatively straightforward way to accommodate a bias against disjunctive categories for these stimuli.

2.3.4.2.2 Domain-General Theories

By contrast, theories that suggest phonetic perception abilities are shared with domain-general abilities struggle with the differences in learning uncovered here. This is most straightforward for direct realism. Under direct realism, both phonetic perception and non-linguistic perception result from the perception of the kinetic attributes of the objects producing the sound and the motor gestures necessary to create the sounds (C. Fowler, 1990). For speech sounds, these are the motor gestures of speech production; for non-speech sounds, these are perhaps the motor actions necessary to make an object make sounds. One’s perception of a saxophone playing, for example, may be enhanced by motor experience producing saxophone sounds (Rizzolatti & Sinigaglia, 2010). However, this leads to an odd conundrum for direct realist theories of perception: how does one perceive sounds that one has not and indeed *cannot* produce using technology currently available, as with the artificial musical instrument continuum used for the present experiment? No one has played a bonetrom or a pettrum, yet listeners have no problem categorizing tokens of each. This, in combination with the differences observed in learning between the non-linguistic and phonetic categories, makes direct realism appear untenable.

General auditory theories also struggle with a difference in learning between the two conditions. Under those theories, speech perception is a subset of auditory perception more generally (Holt & Lotto, 2008). None of the properties of phonetic perception should be unique to the domain of speech sounds. However, the bias seen here, if taken at face value, seems to show a domain-specific constraint on learning that is not shared with another set of auditory categories. However, a good deal more research needs to be done in order to conclusively decide this question one way or another. Although the bias against disjunctive phonetic categories has been replicated with other sets of stimuli besides the fricatives (in a set of experiments not reported here), only one set of non-linguistic auditory categories was used here. It may be that there is something special about the musical instrument sounds that made the difference between disjunctive and non-disjunctive categories disappear. And many of the objections that could be raised by exemplar-only researchers—that the participants in the fricative continuum may have had previous experience with the items in question, or that the fricative items and instrument items were not actually very comparable in structure at all—could also apply for general auditory researchers.

Of course, despite their differing experience with the fricatives used in the present experiment, it is certainly the case that English and German speakers alike have years of experience with phonetic categories in general, even if the English speakers do not have (much) experience with the sounds used in this experiment in particular. The learners' familiarity with speech sound categories may have led them to come into the present task with a prior expectation—perhaps some sort of over-hypothesis (Kemp et al., 2007)—that speech sound categories should not be disjunctive. Again, this seems challenging to accommodate under exemplar-only approaches to category learning, where such biases would need to be instantiated in terms of "baked-in" exemplars belonging to each category. That said, this would significantly weaken the idea of domain-specificity within the present experiment, as the participants' bias against disjunctive categories would result not from something special about speech but rather the typical domain-general auditory refrain: experience.

Such an idea rests on the premise that disjunctive categories are rare in speech. At first blush, this is a reasonable idea; it is hard to come up with many. But there are some categories that may be considered "disjunctive". In American English, the category /t/ can be realized as [t] (a voiceless alveolar stop), [t^h] (an aspirated alveolar stop), [ɾ] (an alveolar flap), [$t^ʔ$] (a glottalized alveolar stop), and even [ʔ] (a glottal stop). The aspirated and glottalized tokens are in some ways the most interesting ones, as they represent the endpoints of a unidimensional continuum of phonation ranging from a lax glottis (breathy voice for aspiration) to a tense glottis (creaky voice for glottalization). Thus, the /t/ category is disjunctive with regard to the dimension of phonation, as tokens with extreme values are all categorized as /t/, while intermediate values are characterized as /d/.   So, despite their rarity, it is possible that disjunctive categories are present for the learners.

One way to get at this question is to use a range of sounds that differ in their perceived "speech-like-ness". Clicks, for example, are consonants that are found productively and frequently in the languages of South Africa, but they are essentially absent in other languages. Indeed, in languages like English, clicks are primarily used for paralinguistic purposes; the sound that is often represented as "tsk-tsk" in American English conventions is a dental click. Studying the acquisition of categories in a less speech-like speech sound continuum, such as ones found in clicks, could provide a window on to what extent these constraints on disjunctive categories are true for every speech sound, even ones not perceived to be speech-like.

3 Adaptation

3.1 Background

Just as variability is important for understanding categorization, so too is it important for understanding adaptation. Variation in the acoustic properties of speech from speaker to speaker likely does not lead listeners to posit entirely new phonetic categories for every new speaker that they run into. A speech system with such a property would be forever in flux, preventing generalization; people would be incapable of picking up on repetitions across speakers. Instead, listeners are capable of adapting to differences in other speakers without changing categories wholesale, picking up on the regularities present in their interlocutor's speech to learn to appropriately categorize other tokens. This is the process of adaptation, covered in detail in the current section.

3.1.1 Speech Adaptation

Adaptation can happen on many levels. Variation can exist from speaker to speaker, but also from dialect to dialect, accent to accent, and even as the result of artificial manipulations (as with helium inhalation). Listeners can track the source of variability—that is, whether the variation comes from an idiosyncratic pronunciation or a consistent feature of a dialect or accent—when processing speech (Kraljic, Brennan, & Samuel, 2008). The recently-proposed ideal adapter model (Kleinschmidt & Jaeger, 2015) is one way to describe how listeners might go about adapting to the speech of others. According to the ideal adapter model, the processes responsible for phonetic adaptation depend on a listener tracking and selecting appropriate statistical distributions for phonetic categories that typify a particular speaker and that speaker's environment. This knowledge about those distributions then allows the listener to assign prior probabilities to the likelihood of certain individual speech tokens. Here, I discuss speech adaptation to two sources of variation: differences in accents and dialects, and differences in rate.

3.1.1.1 Accent Adaptation

Accent adaptation is the most commonly studied facet of speech adaptation. It is not much of a stretch to say that everyone has had an experience listening to the speech of someone speaking a different variety of one's native language, whether those differences are the result of dialect or of having a different native language. Understanding the speech of someone with a different dialect or accent can sometimes be challenging, but, with some level of exposure, it becomes easier. Below, I summarize some of the research into accent adaptation as a window into how and why adaptation happens in general. I largely discuss accent and dialect adaptation interchangeably. It is reasonable to think that the speech of non-native speakers of a language may differ in qualitative ways (or be processed in a qualitatively different way) from native speakers of a different variety of a language. However, the boundaries between the two types of variation are underexplored, perhaps in part because the processing of non-native dialects has been relatively understudied (Floccia, Goslin, Girard, & Konopczynski, 2006). It is also the case that there is substantial variation within the categories of "accent" and "dialect" in the ease of adaptation. Some non-native speakers, for example, whether due to second language proficiency or idiosyncratic factors, are more comprehensible than others (Bent & Bradlow, 2003).

Most research into accent adaptation has relied on one of two methodological approaches: using naturally-produced, accented speech, or creating (often through artificial means) an artificial dialect or accent. Bradlow and Bent (2008) serve as a good example of the naturalistic approach. They played English sentences recorded by either a single or multiple native speakers of Mandarin Chinese to native English speakers, and examined how accurately the native English speakers transcribed the sentences over the course of the experiment. Maye, Aslin, and Tanenhaus (2008) used an artificial accent. They used a speech synthesizer to create an artificial accent or dialect in which front vowels were systematically lowered; that is the [i] vowel (in “seat”) was pronounced as [ɪ] (“sit”), the [ɪ] vowel was pronounced as [ɛ] (“set”), the [ɛ] vowel was pronounced as [æ] (“sat”), and the [æ] vowel was pronounced as [a] (“sot”). They then examined English speakers’ interpretation of the lexical status of words with those shifted vowels.

Both study types have advantages and disadvantages. Artificial dialects are very well-controlled, but may lack some of the validity of real examples of accents. Naturalistic examples, on the other hand, involve sacrificing the control of the items used, which in turn could lead to unexpected findings just as a result of the individual tokens used in the study. That does not necessarily mean that studying systematicity in naturalistic experiments is impossible. Exposure to Spanish-accented English has been used to probe whether certain segments or certain properties of vowels are more affected by adaptation than others (Sidaras, Alexander, & Nygaard, 2009). But it is important to keep in mind the differences between the studies. Adaptation can also be investigated using event-related potentials (ERPs). The Phonological Mapping Negativity (PMN) is a negative-going ERP component that accompanies the presentation of words in accented speech and peaks between 250 and 300 ms after stimulus onset. Interestingly, the strength of the PMN seems to differentiate accented speech and non-native dialects, with

PMN amplitudes increasing relative to unaccented baselines for non-native dialects but decreasing relative to the baseline for accented speech (Goslin, Duffy, & Floccia, 2012).

The paradigms outlined above have led to a fairly well-rounded picture of the processes of accent adaptation. In some respects, accent adaptation can happen very quickly (although it does not always do so). Reaction times to probe words after the auditory presentation of a sentence show a return to a native-like baseline in as quickly as two to four sentences spoken by a non-native speaker (Clarke & Garrett, 2004). The effects of adaptation depend on a listener's familiarity with the variety being spoken; while London-based speakers of Standard English found Glaswegian English (an unfamiliar dialect) to be harder to understand in noise than their native Standard English, Glasgow-based speakers of Glaswegian English were not equally hobbled by Standard English, as they (and most residents of the United Kingdom) had significant exposure to Standard English (Adank, Evans, Stuart-Smith, & Scott, 2009). Adaptation can spread across speaker groups; listeners who heard a variety of foreign-accented speakers (native speakers of Thai, Korean, Hindi, Romanian, and Mandarin) were more accurate in transcribing Slovakian-accented speech than those who were not exposed to foreign-accented speech, despite not actually having had exposure to Slovakian-accented speech before testing (Baese-Berk, Bradlow, & Wright, 2013). And adaptation seems to differentially affect comprehensibility (ease of processing) and intelligibility (ultimate understanding), with comprehensibility often taking much longer and being incomplete even while intelligibility reaches a native-like baseline (Floccia, Butler, Goslin, & Ellis, 2009).

Accent adaptation is not limited to a typical young adult population; people of any age can come into contact with speakers who do not sound like others of their native dialect. A few studies of accent adaptation have examined accent adaptation in older adults, with results being generally inconclusive (Cristia et al., 2012) and perhaps dependent on the cognitive abilities of the older adults (Janse & Adank, 2012). Many more have assessed adaptation in infants and toddlers, where the strength of adaptation seems to be age-dependent. Schmale and Seidl (2009) found that 9-month-olds found it difficult to pick out words that were produced successively by native English and Spanish-accented speakers of English or by multiple speakers of Spanish-accented English, while 13-month-olds were able to detect the similarities between the words despite the phonetic differences characterizing those speakers. A study of American and Jamaican English found that 19-month-olds raised around American English speakers are capable of recognizing words in Jamaican English in a way that 15-month-olds are not (Best, Tyler, Gooding, Orlando, & Quann, 2009). However, that study used only a limited exposure to Jamaican English to trigger adaptation on the part of the toddlers; exposing 15-month-olds to a longer story spoken in a non-native dialect led to successful adaptation in later testing (van Heugten & Johnson, 2014). 19-month-olds also show the ability to adapt to artificial accents when accessing lexical items (White & Aslin, 2011).

These studies do not prove an inability to adapt to other accents or dialects on the part of any group. Indeed, the fact that most of them only used a single talker with an accent or dialect means that in many of these studies the infants' and toddlers' inability to comprehend the speech of the novel talker may have been the result of factors idiosyncratic to that speaker's speaking style, to the extent of differences between the native and non-native speakers of the language, or in the tokens used to expose the children to the new speech variant. It also may be the case that infants and toddlers are capable of adapting to accent variation in the comprehension of familiar words but not, say, when learning new words. These are all factors that may influence the strength of accent adaptation.

3.1.1.2 Rate Adaptation

It is not just differences in accent or dialect that listeners must adapt to; variation exists across a wide range of acoustic properties. Another frequent object of study in the adaptation literature is speech rate. Some people speak at a faster rate than others, and these rate changes lead to systematic variation in the production of nearly every phonetic segment (Crystal & House, 1988). Yet listeners are capable of comprehending speech compressed to as little as 45% of the original duration after a short period of adaptation (Dupoux & Green, 1997), indicating that speech perception can easily accommodate drastic variation in rate. Still, uniform compression of speech leads to different effects on the acoustic signal from naturalistic variation in rate, particularly because in natural fast speech there is predictable variation between segments in, say, the extent to which they are compressed. In the sections below, I review rate adaptation effects, using both naturalistic and artificial manipulation, in segmental perception and in word segmentation.

3.1.1.2.1 Segmental Perception

Speech rate adaptation has been shown to affect the perception of individual segments of speech. Although occasional studies have investigated the influence of context rate on the perception of vowels (Verbrugge, Strange, Shankweiler, & Edman, 1976), most of this discussion has focused on the perception of consonant contrasts. In many of these studies, the duration of adjacent syllables (or even segments within the same syllable) is taken as a proxy for rate, with relatively short syllables being associated with a fast rate and relatively long syllables being associated with a slow rate. Rate adaptation has been observed for consonant voicing distinctions ([p] vs. [b], [t] vs. [d], etc.) and for consonant manner distinctions ([t] vs. [tʃ], [b] vs. [w], etc.), among others (J. L. Miller, 1981).

[b] and [w], for instance, can both be signaled by similar vowel-initial formant transitions, with the difference between them driven by the perception of the duration of that transition. Long formant transitions lead to the perception of a [w], while short transitions lead to the perception of a [b]. However, the perception of those transitions is also modulated by the duration of adjacent segments; relatively long adjacent segments lead to more frequent perception of [b], while relatively short adjacent segments lead to more frequent perception of [w] (J. L. Miller & Liberman, 1979). In other words, speech rate is relative. Speeding up one portion of an utterance leads other portions of the same signal to sound relatively slow by comparison; conversely, slowing down part of an utterance leads other parts of a signal to sound relatively fast. Rate changes are also associated with changes to the internal structure of categories. In voice onset time (VOT) continua, small VOTs are typically perceived as voiced (e.g., [b]), while long VOTs are typically perceived as voiceless (e.g., [p]). However, monotonically increasing VOT can lead to interesting results: tokens that are still perceived as voiceless, but as bad examples of a voiceless category. Which tokens are perceived as “good”, and which as “bad”, are also subject to context rate effects (J. L. Miller & Volaitis, 1989).

Some studies have found that these context effects on segments are surprisingly robust to other changes. For example, in one study that used speeded responses to examine some aspects of moment-to-moment processing, listeners who were forced to give fast voicing judgments reliably behaved as if the syllables in question were short (J. L. Miller & Dexter, 1988). This is a particularly interesting finding because listeners did not merely treat the syllables in question as having an *indeterminate* rate but as having a reliably *fast* rate, suggesting that some aspects of rate processing are obligatory even under amplified task demands.

Rate information also seems to penetrate through signal discontinuities. Consider a very simple word-initial consonant contrast, [bi] ("B") versus [pi] ("P"). Abruptly changing the fundamental frequency information of [i] does not affect its influence on the perception of syllable-initial voicing in the stop contrast, even when such a large change in fundamental frequency accompanies a shift in speaker identity in speech (Sawusch & Newman, 2000). Even rate information from a different (perceived) location and different (perceived) talker from an attended speech stream affects the perception of ambiguous voicing, if only to a very small extent (R. S. Newman & Sawusch, 2009). As mentioned in the introductory chapter, non-human animals also show human-like use of rate cues in voicing distinctions; the contrast between [b] and [w] is also influenced by the duration of the rest of the syllable for budgerigars (Dent et al., 1997)

Such studies have seen some pushback. Diehl, Souther, and Convis (1980) found that the rate effects are strongly context-dependent; for example, they found that using fundamental frequencies that resembled those produced by women led to adaptation effects in the *opposite* direction as those typically observed, although later studies that have used female voices have largely not replicated this reversal. There have also been criticisms of the stimulus parameters used in the perception studies involving VOT and vowel length. In particular, it has been argued that the inverse relationship between VOT and vowel length usually deployed in experiments of rate adaptation are generally not found in production (Kessinger & Blumstein, 1998), while examination of the perceptual effects of naturalistic distributions shows very little effect of rate adaptation on perception (Shinn, Blumstein, & Jongman, 1985).

In contrast to these studies of adjacent vowel durations, the perception of consonantal cues has only sporadically been linked to the perception of far-away (i.e., distal) rate cues. As mentioned earlier, most of these studies use the duration of adjacent syllables or segments as a stand-in for rate. Although it is the case that the duration of adjacent segments often correlates with the rate of the rest of a sentence, this is not always the case. When both distal information and close-by (i.e., proximal) rate information are available, the perception of segmental contrasts seems to be more strongly driven by proximal information than by distal information. Summerfield (1981) looked for the effects of distal timing information on contrasts in word-initial voicing and found almost no evidence for an influence of segmental timing information in segments other than the syllable immediately preceding the initial consonant voicing ambiguity. Although changes to the distal rate around a voicing ambiguity can drive changes in which items are considered to be the best examples of a category, they do not influence the *number* of items that are considered to be good examples of a category (Wayland, Miller, & Volaitis, 1994). Subsequent researchers proposed the idea of a strict temporal window of approximately 400 ms within which timing information can influence a segmental contrast (R. S. Newman & Sawusch, 1996). Reports of successful distal adaptation effects on segmental perception have largely been confined to particular rhythmic contexts, where, for example, only sets of stressed syllables are lengthened (G. R. Kidd, 1989).

3.1.1.2.2 Word Segmentation

In typical orthographic conventions of English, spaces separate words. There is no equally consistent cue for word boundaries in spoken language. As with categorization, the problem of word segmentation—parsing discrete words within a continuous signal—has vexed speech perception researchers, who have tried to explain how listeners can segment speech despite the frequent acoustic ambiguities in the signal. These ambiguities are surprisingly abundant, as the frequency of word-segmentation-related puns suggests. Consider the sign in Figure 15. The creator of the sign was certainly aware that the words “ovary action” and “overreaction” are different words. In spoken language, the two parses are differentiated only by the acoustic properties of the speech sounds used and the context in which the sounds are found; the pun leverages these ambiguities for comic effect. Rate adaptation affects the perception of ambiguous word boundaries in a way reminiscent of some of the stronger studies of segmental perception.

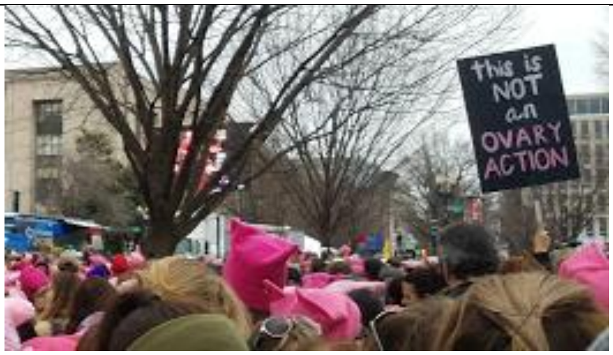


Figure 15. Sign from Women's March on Washington, January 21, 2017

In many models, word segmentation is a consequence, or perhaps even merely a side effect, of word recognition. Under the initial version of the Cohort model (Marslen-Wilson & Welsh, 1978), for example, words are recognized sequentially through parsing speech sounds one-by-one in the signal. A word boundary is posited wherever a complete word can be formed from the string of phonemes being fed into the system. These theories contrast with others that reinforce the primacy of acoustic and probabilistic cues to word boundaries themselves rather than treating word segmentation as the consequence of word recognition. Many more modern formulations of word segmentation models have involved combining the previous proposals related to the sequential nature word recognition to models that involve direct acoustic cues to word boundaries (Davis, Marslen-Wilson, & Gaskell, 2002).

One of the most influential recent proposals was that of Mattys, White, and Melhorn (2005), who found evidence for a hierarchy of cues in word segmentation. In an extensive set of studies, Mattys et al. (2005) compared top-down cues to word boundaries—for example, "lexicality", a desire to parse strings in a way that leads to the perception of valid words of a target language—with bottom-up cues such as coarticulation or word stress. Using a combination of priming and word monitoring tasks, Mattys et al. (2005) found that top-down cues tended to dominate bottom-up cues when available. The authors used their results to posit a hierarchy of word segmentation cues, with lexical ones being most important, acoustic/phonetic cues being secondary, and word stress cues being of tertiary importance. Each less important cue was used only when higher-ranked cues were unavailable.

As with studies of segment contrasts, one issue of interest in the domain of word segmentation has been the use of timing information to segment signals. The timing and duration of acoustic events within the speech signal has been shown to be important across a variety of different phonetic domains. Timing information may trade off with predictability in the speech signal to aid listeners in perceiving sentences (Turk & Shattuck-Hufnagel, 2014). Indeed, as Mattys (1997) pointed out, many of the contrasts between the word segmentation theories discussed above critically hinge on the time course under which the material in the signal is processed. That is, the theories can be differentiated by whether the information is integrated in a strictly linear fashion or whether there is some sort of "buffer" under within which word recognition may take place. Timing seems to be intimately connected to the problem of word segmentation. Turk and Shattuck-Hufnagel (2000) examined the production of triads of lexical items with acoustic ambiguity to word segmentation—for example, *tune acquire*, *tuna choir*, and *tune a choir*—and found evidence for a panoply of word-segmentation-related effects on duration, including lengthening at the beginning of words and lengthening of stressed syllables, which could then be exploited by the listener in order to segment speech.

This has led some researchers to begin exploring the influence of timing information—especially distal speech rate information—on the segmentation of words.

Dilley and McAuley (2008), for example, studied lexically ambiguous syllabic sequences such as *chocolate lyric down town ship wreck*, with the last four syllables capable of being parsed into *downtown shipwreck* and *down township wreck*. For these syllables, the proximal context was considered to be the last three syllables (*town ship wreck*), while the distal context was defined as the immediately preceding syllable (*down*). Dilley and McAuley (2008) found that just by slowing the rate of the distal context, participants went from parsing the phrase as *downtown shipwreck* to parsing it as *down township wreck*, as the duration of *down* came to be perceived as sufficiently long to support its perception as an individual word.

Although these distal rate adaptation effects were originally demonstrated in fairly artificial contexts, they have since been extended to more naturalistic ones. Distal speech rate has been shown to affect the perception of acoustically ambiguous function words, such as *or*, *her*, and *a*. In casual speech, these words are often produced as reduced; that is, short and acoustically indeterminate. They are thus frequently subject to ambiguity in speech, including in the duration and location of word boundaries that set them off from the context (Pluymaekers, Ernestus, & Baayen, 2005). For example, in the phrase *The value went up after her rich neighbors*, the word *her*, if realized as [ɚ] (*'er*), can often blend in with the [ɚ] phone at the end of the word *after*. This creates ambiguity to the existence of the word boundary signaling the word *her*; the listener needs to decide if there is a long enough [ɚ] to sustain the perception of two [ɚ] sounds between *aft*- and *rich* (and, thus, to segment the phrase as *after her rich*), or whether the [ɚ] is only long enough to sustain a single [ɚ] sound (and, thus, only capable of being segmented as *after rich*).

Dilley and Pitt (2010) studied sentence fragments with these ambiguities and other analogous ones, examining whether rate adaptation would affect the perception of these acoustically ambiguous regions within the sentences. They defined the distal context as anything more than a syllable removed from the point of the function word ambiguity in the sentence. In the case of the *rich neighbors* sentence, this was *the value went up af-* and *-ich neighbors*. The proximal context was the critical region of the sentence: *ter (her) r-*. They found that, just by changing the rate of speech within the distal context region, listeners' perception of the critical word segmentation ambiguity changed. In particular, slowing down the distal context led participants to perceive one less word boundary within the critical target region, perceiving *after rich* rather than *after her rich*. The effect sizes were not small. In unmodified versions of the sentence fragments, participants perceived the function word almost 80% of the time; with a slowed distal context (but without any manipulation of the proximal context whatsoever), however, this rate dropped to approximately 30%. These results were later extended to show that the average speech rate across an entire experiment can also influence the perception of individual sentences within the experiment, with relatively slow experimental contexts leading listeners to perceive each individual distal context as slower, thus depressing function word report rates (Baese-Berk et al., 2014).

These effects are appreciably stronger than other cues known to influence word segmentation. In one set of studies, Heffner, Dilley, McAuley, and Pitt (2013) found that distal speech rate more strongly influenced word segmentation than did a set of cues known to induce the perception of word boundaries: changes in intensity and fundamental frequency around the erstwhile boundary and changes in the duration of the ambiguous vowel in question. Intriguingly, given the predictions of Mattys et al. (2005), distal speech rate has been shown to play a greater role than certain top-down cues as well. In a study that examined ambiguous syllable sequences similar to those in Dilley and McAuley (2008), Dilley, Mattys, and Vinke (2010) found that the distal prosodic patterns induced within the context of the ambiguous syllabic sequences had stronger effects on segmentation than did semantic priming from the initial, unambiguous sequences. Similarly, the same ambiguous sentences used in Dilley and Pitt (2010) also allowed for a comparison of the effects of distal rate to the grammatical structure of the critically ambiguous function words within the sentence. Again, it appeared that distal speech rate proved a stronger cue to word segmentation than did the top-down cue of grammatical structure (Morrill, Baese-Berk, Heffner, & Dilley, 2015).

As with accent adaptation, aging has also been studied for rate adaptation. Older adults perceive timing information differently from younger ones (Craik & Hay, 1999). In general, older adults prefer events to be timed more slowly (McAuley, Jones, Holub, Johnston, & Miller, 2006). In one study I performed earlier during my time at UMD, I examined whether older adults might systematically differ from younger adults in their use of distal speech rate cues to word segmentation. Older adults, for example, tend to have problems understanding artificially compressed speech (Gordon-Salant & Fitzgibbons, 2001). I expected older adults (aged 55-65 in this study) to use distal rate less than younger adults in resolving word segmentation ambiguities, and to compensate for this by using top-down cues more. But that was not what was found. Instead, older adults and younger adults used both cues to almost exactly the same extent. This suggested that the differences that have been observed in the perception of speech rate timing did not necessarily entail differences in the use of timing information for real-world tasks. Perhaps the older adults were able to exploit the relative contrasts between distal and proximal speech rates in a way that was identical to younger adults (Heffner, Newman, Dilley, & Idsardi, 2015).

The general pattern, then, is for strong rate adaptation effects in segmentation contrasts, but weak rate adaptation effects for segmental contrasts. However, the distal rate adaptation literatures differ in a considerable number of other ways, including in methodologies. I recently performed an experiment that attempted to iron out all of these differences, using methods and contexts as similar to each other as possible when testing the differences between distal timing information use in segmentation and segments. After holding as much else as possible constant, I found that the support for rate adaptation effects depended strongly on the ambiguities being assessed as well as the definition of "distal" that was adopted. Although segmentation ambiguities (such as *Canadian oats* versus *Canadian notes,* which differ in the location of the word segmentation boundary near the [n]) showed stronger evidence of distal adaptation effects than segmental ambiguities (such as *Canadian coats* versus *Canadian goats*, which differ in the voicing specifications of the word-initial consonant) for word-initial contrasts, adaptation effects were strong for word-final contrasts for both segmentation (*bee knowledge* vs. *bean knowledge*) and segments (*beet knowledge* vs. *bead knowledge*). Furthermore, adaptation effects were relatively weak when the definition of "distal" excluded information within 400ms of a potential word boundary, indicating that at least part of the differences observed between the studies were the result of differences in the definition of "distal". This was true both for participants at UMD and participants recruited from Amazon’s Mechanical Turk crowdsourcing framework (Heffner et al., 2017).

3.1.2 Event Segmentation

The idea that it is necessary to break a continuous stream of input into discrete chunks is not unique to speech perception; it is also seen in the domain of action perception (Zacks & Tversky, 2001). Many of the same ambiguities that are faced by listeners perceiving sentences are also evident in visual perception. Consider learning a new dance for the first time. In order to learn the dance moves, it is necessary to segment the dance into a series of steps. But the learner faces a conundrum: what actions should be considered one step rather than two? The points at which one step begins and the next ends in fluid dance are just as opaque to the casual observer as the points at which words begin and end in fluent speech. Except perhaps in an unusual pedagogical situation, there is no one on the sidelines holding up a flag to indicate when one step gives way to another. The learner then must rely on context cues and her knowledge of the kinematics of bodies in order to assemble the dance into a series of steps. The process of parsing actions into segments is known as event segmentation (Tversky & Zacks, 2013; Zacks & Swallow, 2007). Although outside the scope of the present dissertation, the segmentation of action sequences into events has also been discussed in the semantics literature, as languages require that continuous actions are described in terms of discrete words (Bohnemeyer et al., 2007).

Unlike the segmentation of speech, event segmentation has some terminological challenges; namely, there are many possible timescales that might represent a single “event”, while categories such as “segment”, “syllable”, “word’, “phrase”, “sentence”, and so on have better-defined (if still sometimes ambiguous) boundaries. To use the “learning a new dance” routine as an example, the entire episode of “dancing” might be perceived as an event when considered at the timescale of a day, but is too broad of a useful description when considered at the timescale of a single minute. Quite a bit of research has focused on differentiating fine-grained versus coarse-grained perception of actions in segmentation (Newtson, 1973). These experiments often involve participants watching a movie and pressing a button on a keyboard corresponding to a time at which they perceive one action ending and the next beginning. Most modern examinations of event segmentation have suggested that events are structured hierarchically, in a way reminiscent of hierarchical prosodic structures (Shattuck-Hufnagel & Turk, 1996), with smaller, non-overlapping events combining piece-by-piece to make up larger ones (Tversky & Zacks, 2013; Zacks & Swallow, 2007; Zacks & Tversky, 2001).

As with word segmentation (Mattys et al., 2005), event segmentation is the result of a combination of top-down and bottom-up cues. Bottom-up cues to event segmentation include movement cues (Zacks, Kumar, Abrams, & Mehta, 2009), while top-down cues involve things like experimental instructions (i.e., telling people to focus on fine-grained or coarse timing information). Event boundaries are sometimes associated with “breakpoints”, seemingly invariant points within even scrambled slideshows that exhibit a great deal of physical change from one moment to the next (Bridgette Martin Hard, Recchia, & Tversky, 2011). They can also be prompted by the appearance of new objects in the visual scene (Tauzin, 2015). People are capable of tailoring the cues they use for event segmentation according to the situation they are in;

for example, when perceiving social event boundaries, participants often focus on eye and face information (Boggia & Ristic, 2015).

These types of information often trade off against each other. The salience of bottom-up movement features is increased when participants are asked to segment events in a coarse-grain fashion, indicating that explicit top-down instructions can actually increase reliance on bottom-up cues (Zacks, 2004). Meanwhile, an understanding that the actions being viewed do not serve a direct functional purpose (as in a religious ritual) encourages fine-grained event segmentation (Nielbo & Sørensen, 2011). Even actions that are unfamiliar, played backwards to viewers, or visually inverted are often segmented in a way similar to how they are segmented with top-down knowledge, suggesting that top-down cues may be less effective in event segmentation than in word segmentation (Bridgette M Hard, Tversky, & Lang, 2006; Hemeren & Thill, 2011).

Neural studies have also been used to examine event segmentation. Event boundaries are also generally correlated with functional magnetic resonance imaging (fMRI) activation in frontal and occipital cortex, especially visual area MT (also known as V5), an area that is said to be associated with object motion (Zacks, Speer, Swallow, Braver, & Reynolds, 2007). This activation is accompanied by additional activation in frontal and parietal regions when those actions are perceived to be *meaningful*, rather than ones that were generally opaque to viewers, as in actions taken in tai chi (Schubotz, Korb, Schiffer, Stadler, & von Cramon, 2012). In the neuropsychological literature, traumatic brain injury (TBI) in a military population was associated with poorer processing and understanding of events, especially fine segmentation, when compared to an unimpaired military population (Zacks, Kurby, Landazabal, Krueger, & Grafman, 2016).

One frequent topic of discussion in the event segmentation literature relates to the encoding and retrieval of events in memory. The hierarchies that listeners assemble when segmenting events show up again when asked to recall them later (Zacks & Tversky, 2001). In one study, in which participants had to perform a recognition task after segmenting an ambiguous action, participants were more likely to recognize objects that were present at event boundaries (Swallow, Zacks, & Abrams, 2009). By and large, it seems to be the case that memory representations, both short-term and long-term, are updated at event boundaries (Kurby & Zacks, 2008). Event memory is much more strongly associated with event segmentation success than it is with domain-general memory-related abilities, such as working memory (Sargent et al., 2013). Making event boundaries extremely salient—through a combination of a bell sound, the visual presentation of a large red arrow, and a brief pause in the action on the screen—also seems to enhance memory for a visual scene more than putting such cues in the middle of an event does (Gold, Zacks, & Flores, 2017). Interestingly, older adults may be less hierarchical in their processing of events than younger adults, a fact that appears to be connected to their impaired event recall (Kurby & Zacks, 2011).

Event segmentation is often tied to studies of narrative in movies and books. Written narratives often use elements such as temporal markers (“next week”, “the very next day”, etc.) to signal shifts within stories; these temporal markers lead to event segmentation, which in turn affect readers’ memory for elements of the story in question (Speer & Zacks, 2005). This, in turn, is associated with a widely distributed array of activations in fMRI, including medial temporal gyrus (Ezzyat & Davachi, 2011). Narratives in film have also been the object of study in the event segmentation literature. Many studies have examined the ways that narrative shifts and different types of cuts can lead to scenes being processed as separate events or as single actions (Cutting, 2014), with even discontinuities in space and time being insufficient to break action apart into separate events (Magliano & Zacks, 2011). Many of the properties that lead viewers to posit an event boundary are correlated with the visual properties of the stimulus, without recourse to the top-down goals of the actors in the film (Cutting, Brunick, & Candan, 2012).

Only a handful of studies have explicitly looked at the effects of timing information at all, let alone adaptation, on event segmentation. Some of these investigations have centered on long-scale time representations. For example, actions that are weeks, months, or years in the future are often conceived of in ways that are more abstract and more goal-directed than actions that are perceived on shorter timescales (Trope & Liberman, 2003). Other studies have investigated whether the temporal overlap of different events aids in their segmentation, with event endings seemingly being more powerful than the beginnings of other events in determining the timing of event boundaries (Lu, Harter, & Graesser, 2009). Still, almost no studies have examined the role of duration information per se on event segmentation, nor the rate context surrounding an action, as I and others have done with speech perception (Dilley & Pitt, 2010; Heffner et al., 2013; Reinisch, Jesse, & McQueen, 2011).

Interestingly, despite the paucity of studies related to the influence of duration on event segmentation, the effects of event segmentation on (perceived) duration have been examined. Faber and Gennari (2015) used animations of geometric shapes that varied in the number and composition (specifically, the between-event similarity) of the perceived events within the animation, as determined by item ratings. They found that animations that were segmented into a larger number of actions, and animations in which the actions that occurred were seen as being dissimilar from one another, were perceived to last longer than animations with fewer or more similar actions, and took longer to mentally simulate. Similar effects were found for studies of prospective duration, with estimates of the duration of future actions being predictable in part from the structure of the actions yet to come (Faber & Gennari, 2017).

3.1.3 Event Segmentation and Word Segmentation: Parallels

Given the similarities between the literatures above, the time may be right to attempt to find crossovers between word segmentation and event segmentation. Such thoughts have been stated before. Both word segmentation (Bion, Benavides-Varela, & Nespor, 2011) and music perception (Pearce, Müllensiefen, & Wiggins, 2010; Sridharan, Levitin, Chafe, Berger, & Menon, 2007) have been directly compared to event segmentation.   Peña, Bion, and Nespor (2011) examined the iambic-trochaic law, initially posited in the auditory realm, and applied it to event segmentation. The iambic-trochaic law refers to the observation that two consecutive events that differ in duration tend to be segmented with a short unit followed by a long unit, while consecutive events that differ in pitch or intensity tend to be segmented with a loud or high-pitched unit followed by a soft or low-pitched unit. They found that similar principles are in effect for visual event segmentation as well; for example, events with a long final sub-component tend to be better remembered than ones with a short final sub-component.

Perhaps the best-researched analogue between the two types of segmentation has come from studies of transitional probabilities. It has been claimed that recurring statistical patterns within action sequences lead to structure-building, with less-predictable actions generally being perceived as the onset of a new event (Reynolds, Zacks, & Braver, 2007), just as less-predictable segments or syllables are frequently associated with word boundaries (J. R. Saffran, Aslin, & Newport, 1996). This has led to investigations of the role of transitional probabilities and statistical learning in action perception. Endress and Wood (2011) exposed participants to a series of animations formed from the concatenation of three separate action elements of the form AXB, with the A and B actions maintaining a recurrent and predictable relationship (i.e., the presentation of any particular action A *always* preceded the presentation of a single, different action B). After a period of familiarization, the participants were then above chance at discriminating a sequence following the form AXB than one of the form BAX or XBA (Endress & Wood, 2011). Statistical learning of common action sequences is also present for patterns of moving dots on a screen, with exposure to frequent patterns of motion during familiarization leading to later discrimination of those frequent patterns from ones that never appeared (Ongchoco, Uddenberg, & Chun, 2016).

Other parallels come from the developmental literature. Given the importance of segmentation to speech perception and to event perception, it is perhaps unsurprising that children quickly and effectively learn to segment the world. This has been demonstrated both for speech and for events, but perhaps most robustly in speech perception. Infants as young as seven-and-a-half months old can extract words from the context of a sentence and distinguish them from other words (Jusczyk & Aslin, 1995), and at the age of six months use familiar words (e.g., *mommy*, their name) in order to learn where adjacent words begin and end (Bortfeld, Morgan, Golinkoff, & Rathbun, 2005). In a preferential looking paradigm, 8-month-old infants preferred to orient towards speech samples with pauses placed at phrasal boundaries over speech samples with pauses placed at non-boundary locations (Hirsh-Pasek et al., 1987); later studies with 9-month-olds indicated that this preference persisted even with adult-directed and low-pass-filtered speech samples (Jusczyk et al., 1992). Studies with 11-month-olds showed that this preference for pauses inserted at phrase boundaries also extended to the word level, with the 11-month-olds looking longer at utterances with pauses placed at word boundaries than utterances with pauses placed in the middle of words (J. Myers et al., 1996). Infant word segmentation abilities are important. 12-month-olds with better word segmentation skills showed larger expressive vocabularies at 24 months (R. S. Newman, Ratner, Jusczyk, Jusczyk, & Dow, 2006), suggesting that successful segmentation at a young age has benefits later in development.

It has been abundantly demonstrated that infants are able to use statistical regularities in syllable-to-syllable co-occurrence in order to segment words. Consider the phrase *happy baby*. Given the syllable [hæ] in *happy*, it is probably quite likely to be followed by the syllable [pi], as the word *happy* is a common one in child-directed speech. However, the transitional probability of the syllable [bej] in *baby* given the previous syllable [pi] is much smaller than that of [pi] given [hæ], as the word *happy* can often be followed by a wide variety of possible words (e.g., *birthday*), many of which have different initial syllables. Learning to take advantage of these cues is a process known as statistical learning, which eight-month-olds can do even within the context of a short artificial language experiment (J. R. Saffran, Aslin, et al., 1996). Statistical learning seems to be stronger when the voice used is speaking using infant-directed speech (Thiessen, Hill, & Saffran, 2005), and can quickly be exploited for the sake of learning new words (Graf Estes, Evans, Alibali, & Saffran, 2007).

What other cues are infants and young children able to exploit in order to segment words? Many of the experiments testing this have compared the strength of different acoustic cues to the statistical regularities in speech discussed just previously. For nine-month-olds, stress can affect word segmentation even more than statistical cues can (Mattys, Jusczyk, Luce, & Morgan, 1999; Thiessen & Saffran, 2003), and seems to be assimilated into infants' lexical representations of individual words (Curtin, Mintz, & Christiansen, 2005), although evidence is more mixed at seven months of age (Curtin et al., 2005; Thiessen & Saffran, 2003). Eight-month-olds can use coarticulation and stress more than statistical cues to segment speech (Johnson & Jusczyk, 2001). Nine-month-olds can exploit the different realizations of phonemes, depending on their position within a word (i.e., allophones, as in the aspirated [t^h] of *top* compared to the unaspirated [t] of *stop*), when provided in conjunction to other statistical cues, and by ten-and-a-half months can use them independently of statistical cues (Jusczyk, Hohne, & Bauman, 1999).

Infants and toddlers are also capable of segmenting the world around them into events. Five-and-a-half month olds maintain memories of simple events more strongly than they did memories of individual faces or objects involved in doing the events across the course of a few hours and a few weeks (Bahrick, Gogate, & Ruiz, 2002). As with speech, six- and eight-month-old infants are capable of segmenting individual actions from an event stream, showing analogous segmentation abilities at a similar age (Hespos, Saylor, & Grossman, 2009). These abilities index events specifically rather than just smaller chunks of the familiarization videos, as replaying inter-event transitions led to no increase in looking over a baseline (Hespos, Grossman, & Saylor, 2010; Hespos et al., 2009). Event boundary placement seems to be a skill that is fairly early to master. Ten-month-olds familiarized with a video of a simple action looked more to repetitions of that video with pauses located away from naturalistic event boundaries than repetitions with pauses located at an event boundary (Baldwin, Baird, Saylor, & Clark, 2001). Interestingly, the directionality of this effect is the *opposite* of that found for word boundaries in speech, as infants preferred listening to utterances with pauses at word and phrase boundaries rather than away from those boundaries as in event segmentation. Conversely, similarly-aged infants preferred scenes with non-linguistic acoustic tones playing in sync with event boundaries to ones with tones scattered randomly (Saylor, Baldwin, Baird, & LaBounty, 2007).

As with the adult literature, the developmental literature has begun examining parallels between word segmentation and event segmentation. Perhaps unsurprisingly, this has often taken the form of studies of statistical learning for event segmentation, with a few studies showing that infants are capable of using statistical information to parse a stream of actions into individual events (Baldwin, Andersson, Saffran, & Meyer, 2008; Roseberry, Richie, Hirsh-Pasek, Golinkoff, & Shipley, 2011; Stahl, Romberg, Roseberry, Golinkoff, & Hirsh-Pasek, 2014). Across these studies, then, there are points of overlap in the age of acquisition of each of these domains of language. These similarities in the developmental trajectories of segmentation abilities suggest that other parallels may exist between segmentation both in and out of language.

3.1.4 Summary

Adaptation is a strong and pervasive object of study for speech perception. Although much of the focus in the speech adaptation literature has been on adaptation to accents and dialects, adaptation to rate has been another focus. Despite traditional findings suggesting that the effects of rate adaptation are rather weak on the perception of individual segments, more recent research on rate adaptation in word segmentation has shown much stronger effects. Outside of language, segmentation is also observed in the perception of individual events. Event segmentation and word segmentation share several important properties, and a tiny but growing number of studies have started to investigate parallels between the domains, including in the use of statistical learning to event segmentation. Yet the use of timing cues and the influence of rate adaptation on event segmentation is, to my knowledge, entirely unexplored. In the chapters below, I investigate rate adaptation in speech perception (in particular, the perception of geminate, or doubled, consonants in Arabic) and in event perception (in particular, the perception of actions being performed on a touchscreen), exploring the extent to which rate adaptation can influence each modality.

3.2 Rate Adaptation in Speech

Duration is not just important to segmentation; it influences the perception of a wide variety of phonetic phenomena. The clearest example of this effect is seen through contrastive segment length. In numerous languages, including Italian, Japanese, Finnish, and Arabic (the subject of the present investigation), words can be distinguished merely by the duration of single segments within them. For example, in Italian, the words *beve,* 'he drinks', and *bevve*, 'he drank', are differentiated only by the duration of the medial [v][3]. By contrast, although analogous situations can sometimes arise in English across word boundaries, consonant length alone does not distinguish individual lexical items.

The presence or lack of distinguishable consonant length varies across languages. In languages that do contrast the length of consonants, there are typically two strongly-overlapping cues that inform the perception of the consonants: the absolute length of the consonant, and the relative length of the consonant to the previous vowel. Geminate consonants are, unsurprisingly, long in duration when compared to their singleton counterparts (Idemaru & Guion, 2008; Pind, 1995). On top of that, though, consonant length often shows a robust cue-trading relationship with the length of the previous vowel, with longer consonants often being associated with shorter preceding vowels (Esposito & Di Benedetto, 1999). Yet this particular relationship seems to be subject to some cross-linguistic variability, as seen with Japanese geminate consonants, for instance, which are said to often be preceded by *longer* vowels than singleton consonants (Kingston, Kawahara, Chambless, Mash, & Brenner-Alsop, 2009). Despite the evidence

[3] For purposes of the present experiment, long or geminate consonants will be transcribed as [:], making the transcription of "he drank" [bev:e]. Short or singleton consonants will lack a [:].

for relative cue weighting in production, other authors have suggested that those cues may not translate over to perception (Hankamer, Lahiri, & Koreman, 1989). At least for Japanese speakers, there is a great deal of variability between individuals in their use of absolute and relative duration in the perception of consonant length contrasts (Idemaru, Holt, & Seltman, 2012).

Even in languages with segmental length contrasts, however, it is not solely from phonological or phonetic factors that the duration of segments can vary. The length of the vowels that precede geminate consonants may also vary according to speech rate. Thus, listeners face a dilemma resolving this type of input: is a vowel before a consonant short because the speaker is talking rapidly, or is it short due to the following long consonant, as in the case of Italian and most languages with such contrasts? Although relatively rare, studies investigating the perception of geminate consonants in Icelandic (Pind, 1995) and Italian (E. R. Pickett, Blumstein, & Burton, 1999) indicate that the perception of those consonants is strongly influenced by the length of adjacent syllables, not just vowels, which are often taken as a proxy of the adjacent or proximal rate context. As a result, a consonant with a constant duration might be perceived as either long (i.e., as a geminate) when preceded by a relatively short vowel or as short (i.e., as a singleton) when preceded by a relatively long vowel. While the length of adjacent syllables can often show variation in line with speech rate, ratios of consonants to vowels stay relatively stable even in the face of rate changes (Idemaru & Guion-Anderson, 2010). These findings suggest that listeners can show rate adaptation effects even in the case of consonant length contrasts that are primarily cued by the same durational properties that signal rate changes.

Although the vast majority of the distal rate adaptation literature has taken place in the context of studies of English-speaking listeners' perception of segmentation and segments, there have been a handful of studies that assessed distal rate adaptation in languages other than English. These studies have primarily been performed in cases where the critical ambiguities involved segmentation or the all-or-nothing perception of individual segments rather than for differences in consonant length. In one study using Dutch (Reinisch et al., 2011), for example, listeners heard sentences with phrases ambiguous to the location of a word boundary, such as *eens (s)peer*, 'once spear/pear'. The pairs were ambiguous to whether two [s] sounds abutted the word boundary or whether there was just a single [s] found to one side of the boundary. Similar to findings from English, the rate of speech more than a syllable removed from the potential word boundary influenced the perception of the word boundary, with a slower distal rate leading people to report the doubled [s] less often. Russian speakers were less likely to report a grab bag of acoustically de-emphasized segments and syllables—including function words that were signaled only by the presence or location of a word boundary—with a slower distal rate (Dilley, Morrill, & Banzina, 2013). While the distal rate effects varied from item to item and context to context, they were generally of a scale seen in studies of English word segmentation effects. In the present study, I extend these previous results seen with regard to word segmentation to see the effects of rate adaptation effects on consonant length contrasts.

3.2.1 Experiment 4: Rate Adaptation in Arabic

Words in Modern Standard Arabic (MSA) may vary only in terms of the length of individual segments within a word. Words in MSA are composed of abstract roots and word-patterns. Thus, a single root, such as DRS, may be incorporated to different word-patterns, which carry the morphological information, leading to more than twenty-five different words. Variation between two word-patterns employing the same root may be limited to the varying length of individual segments within the word. The words *darasa*, ‘he studied’, and *darrasa*, ‘he taught’, are differentiated only by the length of the medial [r] sound. As can be seen just by the example of ‘studied’ versus ‘taught’, these length contrasts have urgent importance to the grammatical structure of Arabic; differences between closely related verb meanings, as well as between singular and plural nouns, are signaled through differences in consonant length.

This makes Arabic an interesting case study, as future experiments could assess the relative importance of content-based cues to word identity and the rate cues to consonant length. As with all languages that have contrasting length specifications, this distinction is strongly cued by the length of the consonant itself (Obrecht, 1965). Arabic consonant length contrasts are also associated with cue trading relationships that resemble those in Italian and other languages. Vowels that precede geminate consonants are relatively short compared to vowels that precede singleton consonants (F. Y. Al-Tamimi, 2004; J. Al-Tamimi & Khattab, 2011). Thus, adaptation is particularly important for Arabic speakers, as listeners must decide whether any particular vowel has a small duration because the vowel precedes a long consonant or because the listener is speaking at a quick rate[4]. In the present chapter, I investigate distal rate adaptation effects

[4] Although outside the scope of the present investigation, Arabic also has contrastive *vowel* length, meaning that the vowel itself may also have a short or long specification.

on Arabic speakers' perception of consonant length contrasts along the lines of my earlier experiments on distal rate effects in segmentation.

3.2.1.1 Participants

20 people participated in this experiment (16 female, 3 male, 1 not stated). All were at least 18 years old (*M* = 27.9[5], Range = 19-50) and had no history of speech or hearing disorders. All were native speakers of Arabic, primarily Peninsular Arabic, and were fluent speakers of Modern Standard Arabic, the standardized variety of Arabic used in writing and in mass media in the Arabic-speaking world. Participants were recruited either in the United States or Saudi Arabia by a native speaker of Arabic. They were compensated at either a $10/hour wage or local equivalent or refused payment. The experiment was performed in line with the guidelines of the University of Maryland, College Park Institutional Review Board (IRB).

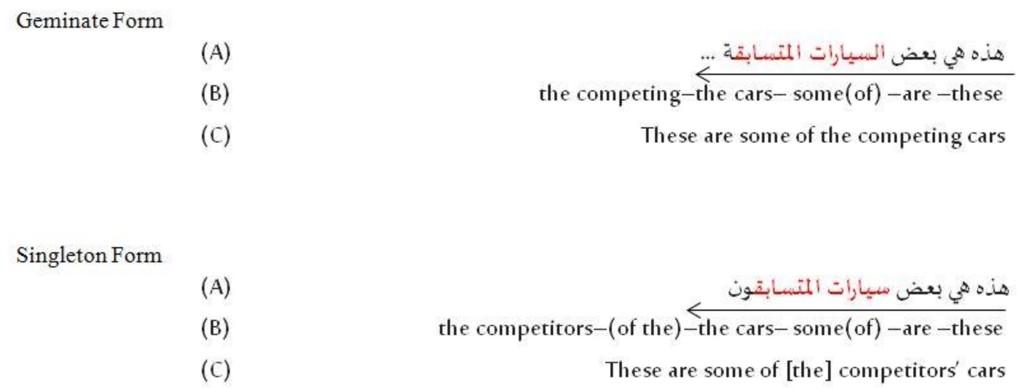
3.2.1.2 Materials

30 sentence pairs were designed with a critical ambiguity in the length of a consonant signaling the presence of a definite clitic. In Arabic, the definite clitic is often transliterated as *al*, and is attached to the beginning of a noun or an adjective that it modifies; *bayt*, 'a house', becomes *al-bayt*, 'the house'. However, two key processes can conspire to render its perception dependent only on the length of a critical consonant. First, the [l] of *al* undergoes complete assimilation to the following consonant if that consonant is coronal. For example, the definite form of the noun *sayaraat*, 'car', is *as-sayaraat*, 'the car'. This makes the length of the consonant phonetically long ([s:]).

[5] Not all participants were comfortable giving their exact age and instead provided an approximate range of ages. For these participants, the midpoint of the range was used as the age for purposes of computing the mean.

Second, when articulated after a word ending with a vowel, the *a* of *al* is usually elided as in the example *ba'du as-sayaraat.* When the word *ba'du* ([baʕd^{ʕ}u]), 'some', precedes *as-sayaraat*, 'the car', the *a* of *as-* is often elided. As such, then, in cases when the determiner clitic is preceded by a vowel-final word and attaches to a word starting with a coronal consonant, the only disambiguating element as to whether that determiner is present is whether the consonant the clitic attaches to is long.

The sentence pairs that were created for this experiment differed only in the length of a consonant that signaled the presence of a determiner. To do this, 30 sentences were constructed that were identical up to the point of a critical consonant, then diverged with regard to the length of that consonant. The items that were recorded without an *al* will be referred to as "singleton" items, while the items with an *al* will be referred to as "geminate" items (see Figure 16 for examples of pairs). The glosses marked (C) are the English translation of (A), while the items marked (B) are the word by word translation of the Arabic sentence for the purpose of reflecting word order in Arabic. Note that Arabic is read from right to left, with the (B) items created accordingly. Although later-occurring grammatical and semantic information that disambiguated the length of the critical consonant was included to aid in the speakers' pronunciation of the materials, this information was cut out of the sentence fragments that were played for the participants. 60 filler sentences with no such ambiguities were also constructed, with those sentences being subject to truncation in a similar way to the experimental items. Four native Arabic speakers—two female speakers of Peninsular Najdi Arabic, and two male speakers of Egyptian Arabic—recorded the stimuli using MSA. Items were selected to roughly balance the number of items used from each of the four speakers while also selecting singleton and geminate items that had similar acoustic properties. Although varieties of spoken Arabic in Saudi Arabia and Egypt do have some phonetic and phonological differences, all varieties of Arabic include the properties that make these items ambiguous, including vowel elision and complete coronal assimilation.



Geminate Form

(A) هذه هي بعض السيارات المتسابقة ...

(B) the competing–the cars– some(of) –are –these

(C) These are some of the competing cars

Singleton Form

(A) هذه هي بعض سيارات المتسابقون

(B) the competitors–(of the)–the cars– some(of) –are –these

(C) These are some of [the] competitors' cars

Figure 16. Experiment 4 stimulus example

The 30 items with the critical ambiguity had their distal speech rates modified. The "distal context" was defined as anything more than a syllable removed from the point of ambiguity, in line with numerous previous studies of these effects in word segmentation (Dilley & Pitt, 2010; Heffner et al., 2013). The "proximal context" was defined as everything within a single syllable of the word boundary. For these items, there were three possible context rates: Normal (with no change to the distal rate), Slow (with a distal context length set to 175% of the unmodified version), and Fast (with a distal context length set to 70% of the original duration). The rate of the filler items was also changed to be Normal, Slow, or Fast, but with rate manipulations that affected entire sentences rather than parts of the sentences.

3.2.1.3 Procedure

The experiment used a 2 (Type: Singleton or Geminate) × 3 (Distal Rate: Normal, Slow, or Fast) design. Items were randomly assigned to one of six lists, where each item was assigned to one of the six combinations of Type and Distal Rate. For half of the lists, one item was inadvertently added to the list twice; the second iteration of that item was removed from further analysis. The order of all experimental and filler items was completely randomized for every participant, with the exception of two filler items used for practice at the beginning of the experiment. All participants completed the study using the same computer and headphone, and were encouraged to ask questions before they start the experiment. PsychoPy was used to run the items, and participants heard the sentences presented one-by-one and were asked to write down the sentences that they had heard. They were allowed to repeat an item up to five times before they wrote the sentence down.

3.2.1.4 Analysis

Participants' transcriptions of each sentence were examined for the presence of the critical determiner clitic *al*. For a few items (less than 5%), it was indeterminate whether the participant had transcribed the determiner given transcription errors near the critical region. These trials were not considered for subsequent analysis. However, for the rest, the presence of the definite clitic was coded as either a 1 if the transcription contained the determiner or a 0 if the transcription did not. Type was coded as a factor, while distal rate was coded as a continuous variable. The scale factors expressing the duration of the distal duration with regard to the unmodified version of the clip were base-2 logarithmically scaled to give the numbers that each rate was coded as: the

Normal rate was coded as 0, the Slow rate as 0.807 (the base-2 logarithm of 1.75), and Fast as -0.515 (the base-2 logarithm of 0.70). This preserved the fact that the three rates were not completely independent of each other; the slow rate was slower than the normal rate, which was in turn slower than the fast rate. Generalized linear mixed-effects models were then implemented in the lme4 package (Bates et al., 2016) to compare participants' tendencies to transcribe the definite clitic across combinations of Type and Distal Rate. A model comparison procedure was used to first identify the most complex random effects structure supported by the data, with help from procedures instantiated within the RePsychLing package (Baayen, Bates, Kliegl, & Vasishth, 2015), and to second determine the fixed effects with a significant impact on participants' transcriptions of the critical region (Bates, Kliegl, Vasishth, & Baayen, 2015).

3.2.1.5 Results

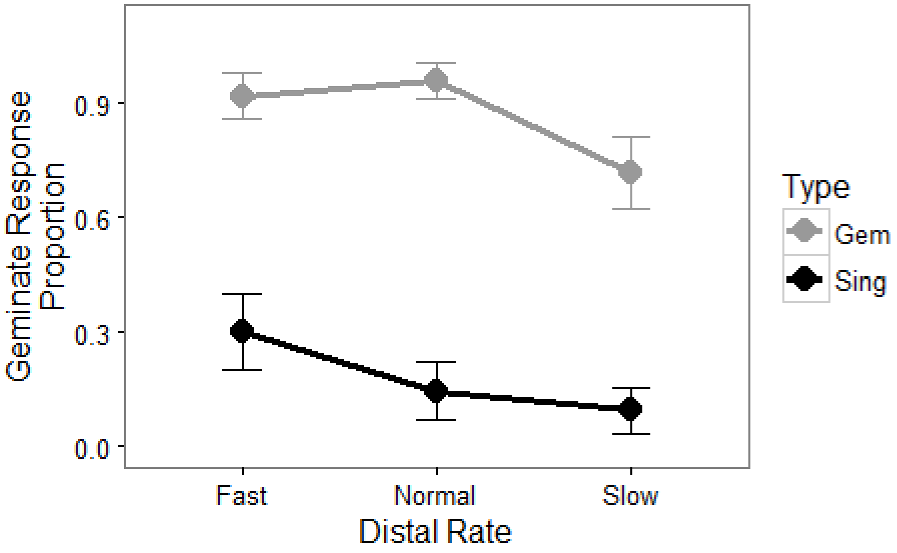


Figure 17. Experiment 4 results

A summary of the results in the experiment is found in Figure 17. Figure 17 illustrates the proportion of critical regions transcribed with a geminate response, by original item type (geminate or singleton) and distal rate (fast, normal, or slow). Error bars are by-participant standard errors. The item type (singular versus geminate) clearly affected the proportion of trials in which participants reported a geminate consonant. However, the slope of each line also indicated support for the idea of distal rate effects on the perception of the critical consonants as well.

These subjective observations were confirmed by modeling. The first step was to assess the ideal random effects structure for the dataset. To do this, an initial model was constructed that had all of the potential fixed and random effects included. The initial model included fixed effects of distal rate and type as well as the interaction between them, random intercepts by participant and by item, random slopes for distal rate by participant and item, and random intercepts for type by participant and item. No correlation parameters were included between random slopes (Bates et al., 2015). Next, to find the maximum number of dimensions supported by the random variation in the data, a principal components analysis (PCA) was performed on the variance-covariance matrix of the model using the RePsychLing package in R (Baayen et al., 2015). This PCA indicated that a maximum of three random components were supported by the variation in the data by item, but only one was supported by participant. Comparing the initial model to an intermediate model with a full random structure by item but only random intercepts by participant showed no significant change in model fit from dropping the random slopes by participant, $\chi^2(2) = 3.18$, $p = .20$. Thus, the intermediate model, with random intercepts by participant and by item and random slopes for distal rate and type by item, appears to provide the most reliable set of random effects.

This intermediate model, then, can be compared to models that lack different *fixed* effects in order to determine the significance of each of these effects. Comparing the intermediate model to one lacking the fixed simple effect of type, as well as the fixed interaction between distal rate and type, would indicate whether type played a significant role in determining participants' perception of the critical consonants. And, indeed, there was a significant decrease in model fit with these fixed effects removed, $\chi^2(2) = 66.0$, $p < .001$. Participants were much more likely to hear the critical consonant as a geminate if it was recorded with that intention, probably in line with the many other acoustic cues present that can indicate the presence of a geminate consonant (Idemaru & Guion, 2008). Comparing the intermediate model to one lacking distal rate (both the simple effect and the interaction with type) also yielded a significant difference, $\chi^2(2) = 15.4$, $p < .001$. Slowing the distal rate made people less likely to hear the critical consonant as a geminate. However, comparing the intermediate model to a final model lacking the interaction term between distal rate and type – yet possessing the effects of distal rate and type independently of one another—showed no significant change in model fit without the interaction, $\chi^2(1) = 0.267$, $p = .61$. The dataset, then, best supports this final model, one that includes distal rate and type as independent fixed factors. Fixed model parameters are available in Table 6, with Geminate as the reference level for the type factor.

Factor(s)	**Estimate (*b*)**	***z***	***p***
Intercept	2.62	8.01	< .001
Distal Rate	-1.90	-3.59	< .001
Type:Singleton	-4.48	-10.6	< .001

Table 6. The best-fitting model in Experiment 4

3.2.2 Discussion

The study described in this chapter tested the effects of distal rate adaptation on the perception of Modern Standard Arabic length contrasts. I predicted that modifying the distal rate by speeding up the context around a singleton consonant in Arabic would lead that consonant to be perceived as a geminate. Slowing down the context, on the other hand, around a geminate consonant in Arabic would lead that consonant to be perceived as a singleton. The findings of the study supported the prediction completely. Fast contexts led singleton consonants to sound relatively long, and thus they were more likely to be perceived as geminates; slow contexts led the geminate consonant to sound relatively short, and thus to be perceived as singletons. This extends findings related to distal rate adaptation to a new set of segmental contrasts used across a variety of languages.

One particularly interesting aspect of the results is that they more closely match the literature on distal rate adaptation effects for word segmentation contrasts, not for segmental contrasts. Examples of such contrasts in English include pairs such as *Minneapolis sale* and *Minneapolis ale*. For these contrasts, distinguishing between the two possible ways to segment the phrase depends critically on the perception of the length of the ambiguous [s] sound. If the sound is long enough to be perceived as two instances of [s], the phrase is perceived as *Minneapolis sale*; if not, it is perceived as *Minneapolis ale*. In each of these instances, context rate appears to strongly influence the perception of the length of this class of consonants (Heffner et al., 2017; J. M. Pickett & Decker, 1960; Reinisch et al., 2011). This can be compared to studies of distal rate adaptation on segmental effects, where adaptation effects are said to be small to non-existent (R. S. Newman & Sawusch, 1996; Summerfield, 1981).

What could explain the difference between the studies of segmental perception and segmentation? In Heffner et al. (2017), we considered four possible explanations. Two of these relied on a qualitative split between segments and segmentation, either in terms of how information is processed or in terms of how information is represented. The two other explanations related to more idiosyncratic differences between the previous studies in the literature, either in terms of what was considered to be "distal" rate context or in terms of the types of items that were used in the previous studies. It is likely that at least some of the differences in effect sizes stem from differences in the region of speech that is described as "distal" between the segmental and segmentation literatures (Heffner et al., 2017). We also judged it reasonable that some of the differences might also be explained by an overreliance in the segmental literature on word-initial voicing contrasts. Almost every study looking at distal rate adaptation effects on segmental contrasts has examined word-initial voicing pairs such as *Minneapolis pail* and *Minneapolis bail*, which are not strongly affected by rate adaptation. Nonetheless, when we looked at word-final voicing pairs such as *beat knowledge* and *bead knowledge*, those pairs were subject to distal rate adaptation effects, ones of an approximately equal magnitude to the effects found in the segmentation literature.

The present studies show another segmental contrast that can be affected by distal rate information: Arabic segmental length contrasts. This provides further evidence in favor of the idea of experimental properties, not a qualitative split in processing or representation, underlying the distinction between segmental and segmentation contrasts. The difference in length studied here clearly falls in the domain of the segment in Arabic; yet it too is subject to distal rate adaptation effects similar to that of segmentation. The vast majority of studies that fail to show distal rate adaptation effects are ones that have involved initial voicing contrasts. In almost all other circumstances, listeners adapt to the rate of far-away information in speech processing. This might occur because voice onset times (VOTs) are perhaps not as rate-dependent as originally thought; the VOTs alone, without recourse to the rate of the surrounding syllables, may provide sufficient information to distinguish between voiced and voiceless tokens (Nakai & Scobbie, 2016). Thus, listeners need more information to distinguish other potentially-ambiguous segmental contrasts, including the Arabic length contrast. More generally, this reinforces the universality of rate adaptation in speech perception. Listeners do adapt to rate information, and do so frequently, across languages and across contrasts. Nonetheless, little is known about rate adaptation in other modalities, which I address next.

3.3 Rate Adaptation in Non-Speech

Given previous findings related to speech rate and phonetic processing, I aimed to see whether similar results could be obtained for event perception, particularly event segmentation. Most studies of event segmentation have relied on naturalistic designs, with unfamiliar real-world actions being the subject of experimental manipulation. Although this approach does have its benefits, particularly in its (relative) ecological validity, it is also more challenging to assess cause-and-effect with regard to individual cues to event segmentation. As such, I created stimuli that are more analogous to the individually-constructed sentences of word segmentation experiments. In particular, I used videos of touchpad interactions that were specially constructed to contain ambiguities in the number of interactions found at the end of an action sequence. These touchpad stimuli were then artificially speeded up or slowed down in a manner analogous to studies examining the effects of distal speech rate on word segmentation. This allowed me not only to see the effects of timing information on event perception but also whether those effects would resemble those found in speech perception.

3.3.1 Experiment 5: Rate Adaptation for Event Perception

3.3.1.1 Participants

45 participants completed the experiment. 4 of those participants were excluded from further analysis: 1 because of a missing demographics survey, 2 due to technical errors, and 1 due to experimenter error. That left 41 participants with analyzable data (15 female, 26 male). All participants were native speakers of English at least 18 years of age (*M* = 20.4, Range = 18-26) and had no history of uncorrectable vision impairments. Participants, recruited from the University of Maryland, College Park community, were

compensated with $5. The experiment usually lasted around 30 minutes, although some participants took up to 45 minutes.

3.3.1.2 Materials

Participants saw a set of 63 experimental videos and 63 filler videos over the course of the experiment, selected from a larger set of videos based on the level of perceived ambiguity in the actions in pilot studies. The videos were recorded using a fixed digital camera at 24 frames per second showing a single, seated actor interacting with a touchscreen device. Each video showed a sequence of 7 or 8 actions: a tap, a press, a drag, a swipe, a double tap, a twist or rotate[6], a pinch, or a spread. The actor was instructed to perform these actions as naturally as possible, although some time pressure was used to ensure that the sequences were reasonably ambiguous. The combination of framing, focus, and camera angle such that the movement of the actor's fingers on the touchscreen was the primary cue available to determine the actions being performed. For example, the actor's head was not present in the frame, and the camera angle prevented the viewer from seeing what, if anything, was happening on the screen (which, in any case, was turned off for the duration of recording). The duration of each video depended on the exact sequence of actions, but generally varied from 5 to 8 seconds in length.

To examine the effects of rate adaptation on the perception of ambiguous events, nearly all experimental items ended in one of three possible sequences of actions: a drag action, a press action, or two tap actions. The drag, press, or (two) tap actions could be

[6] Due to experiment error, "twist" was used to describe this action during the instructions, while "rotate" was the labeled option given during the experiment. However, as this was a "filler" action that was used to pad the sequences with additional possible responses, it is not believed that the competing descriptors affected the final result.

ambiguous in their timing properties with a swipe action, a tap action, and a double tap action, respectively. It was intended that the tap actions and drag actions would be ambiguous to their event segmentation (making it analogous to studies of word segmentation, with generally strong context rate effects), while the press action would be ambiguous to its identity (making it analogous to studies of segment identification, with generally weak rate adaptation effects). The action sequences that ended with two taps could also been seen as ending in a single double tap. Meanwhile, the drag action was intended to be preceded by a press, making the difference between press-and-drag and swipe also a matter of segmentation. However, the properties of the actor’s productions prevented this possibility for press actions, as the “press” and “drag” actions were consistently produced as clearly separate actions, akin to putting a discrete pause between two words in speech. As such, a number of fillers (about one third of the experimental items that ended with a “drag”) were converted into experimental items by cutting the last action out of the clip, leaving the previously penultimate “drag” action as the final action for the purposes of subsequent modifications.

The clips were then rate-modified using free ffmpeg software package, which allows for fine control duration through either dropping or duplicating frames. The critical actions were compressed in duration by setting their duration to be 33% of the original duration (i.e., dropping 2 of every 3 frames), which appeared in pilot testing to be approximately the point of maximum ambiguity with regard to the perception of each action. This was meant to be analogous to the ambiguous items in the speech study; although they were not rate-modified in Chapter 3.2, there was good reason to believe they were at least somewhat ambiguous, whereas I could not be sure of that in the present experiment. It was easier to shorten the drag, press, and tap actions to make them appear to be their “fast” counterparts than vice-versa. Filler actions ended with one of the other actions: twist, spread, pinch, swipe, or double tap.

Besides the manipulation of the critical actions, experimental items also had the rate of the precursor actions within each action sequence modified. This was meant to be analogous to the changes in speech rate that were used in Chapter 3.2. The duration of all but the final action (for drag or press sequences) or all but the last two actions (for sequences that ended in two taps) was modified either by halving it (i.e., dropping every other frame) or doubling it (i.e., duplicating every frame). This represents a more all-encompassing definition of “context” than was adopted in Chapter 3.2, as the previous experiment did not involve the manipulation of the syllable immediately preceding and following the ambiguous consonants. Filler items were modified by a single, uniform scaling factor across the entire action sequence through either doubling its duration, halving it, or leaving it unmodified.

To ensure participants were attending to the entire sequence of actions, rather than just the final actions, a single, 300ms 440 Hz sine-wave tone was inserted somewhere in the clip. For the experimental items, this tone always occurred at the end of the action sequence, just after the last action. For the filler items, it could occur anywhere between the end of the second action and the final clip, although the distribution of tone placements was strongly biased towards the end of the sequence (with a peak at the second-to-last action) so as to not make the distinction between the experimental and filler items obvious.

3.3.1.3 Procedure

The experiment used a 3 (precursor rate: slow, medium, and fast) × 3 (type: tap, drag, and press) design for the experimental items. Participants were assigned to one of three lists, which were counterbalanced for the assignment of each item to each precursor rate. Participants were seated in a sound booth for the duration of the experiment. They were told that they were going to watch action sequences that were silent except for a single tone. Their task was to select one of eight possible responses corresponding to the action that they saw immediately before the tone by pressing a button between 1 through 8 on a keyboard. The tones were played over Sennheiser M40fs headphones. Before the experiment began, the participants were shown examples of each of the eight actions presented in isolation, and they were given plenty of opportunities to ask questions. During the experiment proper, trials (both experimental and filler items) were presented in random order, assigned on a participant-by-participant basis. Each item repeated up to 3 times before participants were expected to respond. During participant responses, a 4 × 2 grid was displayed on the screen that listed the possible responses and their associated keys.

3.3.1.4 Analysis

First, participant responses were coded for accuracy. Tap trials were coded as accurate if the event reported was either a tap or a double tap (92% of trials); drag trials were coded as accurate if the event reported was a drag or a swipe (93%); and press trials were coded as accurate if the event reported was a press or a tap (93%). All other responses were coded as inaccurate and discarded. For the remaining trials, responses were coded as “long response” (and assigned a value of 1) if they were reported as originally recorded, and as a “short response” (and assigned a value of 0) if the short analogue of each original event was reported (i.e., double tap for tap trials, swipe for drag trials, or tap for press trials). Precursor rate was coded as -1 if the rate was fast, 0 if it was normal, and 1 if it was slow, in line with a logarithmic transform of the scale factors applied to the duration of each file. For example, the slow clips involved doubling the duration of the original film (multiplying it by 2), and the base-2 logarithm of 2 is 1.0. As in Experiment 4, mixed models implemented in the lme4 package (Bates et al., 2016) and refined using the RePsychLing package (Baayen et al., 2015) within R (version 3.3.1) were used to analyze the dataset. Model comparison was used to first identify the most complex random effects structure in the data and then to determine the provenance of the fixed effects of precursor rate and type. To aid in model comparison, the BOBYQA algorithm was used to create the mixed models.

3.3.1.5 Results

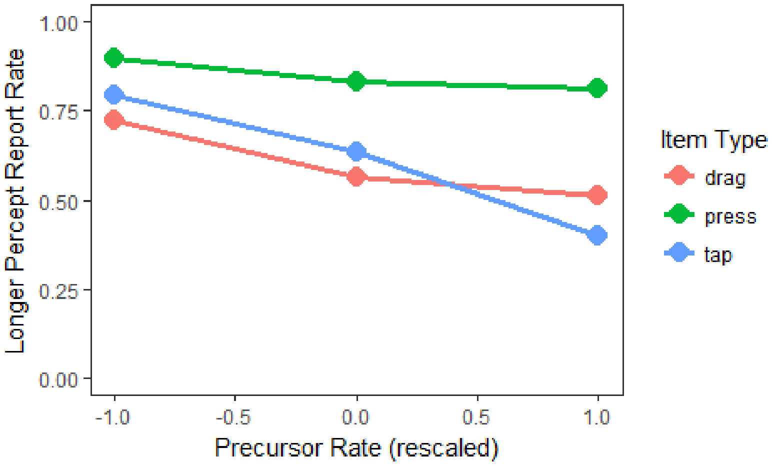


Figure 18. Experiment 5 results

Figure 18 shows the effects of precursor rate and item type on the likelihood that participants perceived a long action (whether a drag, a press, or two separate taps). A value of -1 for the precursor rate represents a context with half of the duration of the unmodified version; a value of 1 represents a context with a doubled duration. Two effects are obvious. First, it is clear that the items used in the study were ambiguous to different extents. For example, press items (with no precursor rate modification, only a modification of the rate of the critical event) were perceived as “press” events about 85% of the time, while drag items were perceived as “drag” events 60% of the time. On top of that, however, the perception of long responses was rate-dependent, particularly for the “tap” actions; people were more likely to perceive the actions as short when the context was relatively long (i.e., slow).

These impressions were confirmed using mixed models. First, the random effects structure of the dataset was computed. An initial model was created to serve as a point of comparison for subsequent analyses of random effects. This model included random intercepts by participant (allowing participants to vary in their baseline propensity to indicate long responses) and random intercepts by item (allowing items to vary in their likelihood of being perceived as long), as well as random slopes for precursor rate by participant (i.e., participants varied in how affected by precursor rate they actually were) and by item (items were also allowed to vary in precursor rate effects), and for item type by participant (participants had different baseline levels of reporting long percepts across each item type). As before, no correlation parameters were included between random slopes. A PCA was performed on the variance-covariance matrix of the model in the RePsychLing package (Baayen et al., 2015). This PCA indicated that the random variation in the model indicated that all of the random components in the model were plausible to include, although model comparison was performed to determine the extent to which they truly contributed variation. In the end, removing any of the three random slopes led to a significant decrease in model fit, with the smallest change in fit coming from the model that lacked the random slope of precursor rate by participant, $\chi^2(1) = 6.20$, $p = .01$. As such, the initial model was used as the point of comparison to determine the provenance of fixed effects.

The effects of precursor rate and item type were first examined by comparing the initial model to models that lacked both the main effects of each factor and the interaction between precursor rate and item time. This comparison between the initial model and the models lacking each individual effect was significant for both the model lacking effects of precursor rate, $\chi^2(3) = 43.6$ $p < .\ 001$, and one lacking effects of item type, $\chi^2(4) = 32.1$, $p < .001$. Significant variation in participant behavior could be explained by the fixed effects of precursor rate and context type when considered independently of each other. Put another way, there were significant differences between item types and precursor rates in the likelihood that people saw the events being depicted as long. Additionally, comparing the initial model to one lacking solely the interaction between precursor rate and item type also yielded a significant effect, $\chi^2(2) = 8.80$, $p = .01$, indicating that some item types were more affected by precursor rate than others. In particular, the “tap” items were more affected by rate than the “press” and “drag” actions. Table 7 shows the fixed model parameters for the best-fitting initial model, with “drag” actions as the reference level for the factor of item type.

Factor(s)	**Estimate (*b*)**	***z***	***p***
Intercept	0.556	1.59	.11
Precursor Rate	-0.666	-4.16	< .001
Type:Press	1.82	4.00	< .001
Type:Tap	0.0999	0.258	.80
PR × T:P	0.224	0.990	.32
PR × T:T	-0.467	-2.22	.03

Table 7. The best-fitting model in Experiment 5

3.3.2 Discussion

I set out to determine if rate adaptation effects resembling those in speech perception could also be found in the perception of events. The answer from the present study is “yes”. The perception of individual events is influenced by the rate that precursor events are seen to be performed at. This appears to occur in much the same way as the perception of individual segments and individual words is affected by the speed that context words are heard to be produced at. Slowing down the rate around two actions that are seen as taps without any modification can turn the perceived action into a double tap. Slowing down the rate around an action that was originally produced as a drag can turn the perceived action into a swipe.

Two aspects of these results should be highlighted. First, although, as the literature review presaged, the initial thought was to examine just event segmentation, it is clear that rate adaptation effects also extended more generally to other aspects of event perception. The difference between a swipe and a drag, at least for the instances of each action performed for this experiment, likely does not have to do with segmentation. Although studies of rate adaptation in the speech domain have found the strongest effects of adaptation on word segmentation (Dilley & McAuley, 2008; Dilley & Pitt, 2010), adaptation effects have also been found for word-final segments (Heffner et al., 2017).

Experiment 4 uncovered analogous effects with regard to the perceived length of Arabic consonants. All of these effects are similar in nature to the non-segmentation-related event perception changes in the present experiment; rate adaptation changed the type of event perceived for some of the actions, not the number of events.

Second, the rate adaptation effects here were not solely distal in nature. I manipulated the duration of every event in the action sequence other than the critical one, rather than, say, only modifying non-adjacent actions within the sequence. This makes the present findings less analogous to my previous studies of rate adaptation in speech. However, subsequent experiments can tease out the extent to which listeners use widely-dispersed rate information in perceiving events.

Third, the current study also does not address what timing information is being tracked in the context. Previous studies in speech have found, for instance, that stressed syllables appear to have a privileged position in establishing rate expectations within a sentence (G. R. Kidd, 1989). Figuring out the relevant units of event perception, on the other hand, is more challenging to assess, and awaits further study.

3.3.2.1 Relevance to Event Perception

This project advances research into event perception in a couple of directions. The first relates to the importance of rate to the perception of events, including to the identification and segmentation of different events. Most studies of bottom-up cues to event boundaries have focused on attributes of event boundaries such as changes in the motion of objects in the visual frame (Zacks et al., 2009) or physical change across frames (Bridgette Martin Hard et al., 2011; Tauzin, 2015). These cues can be considered equivalent to, say, frequency cues to segment identity or word boundaries in speech.

However, timing information is also important to speech; the length of segments can also influence where a word boundary is perceived (J. M. Pickett & Decker, 1960). This experiment demonstrates that timing information is also critical to the perception of events, both in terms of segmentation (as with two taps versus a double tap event) and in terms of differentiation between different possible events (as with a swipe versus a drag event). This observation suggests a new dimension of possible research in the event perception literature: time.

However, it is not just raw timing that was manipulated in the present experiment. I also probed adaptation, the idea that listeners would adjust to the rate context of actions in order to perceive them. This is why precursor rates were manipulated *around* ambiguous events. In this case, the rate context modified was every action before an ambiguous event (or set of events); in contrast to Experiment 4, then, this was not strictly a distal modification, but also included timing information throughout the entire action sequence, up to the final, ambiguous action. Still, this suggests that listeners adapt to the rate of action sequences in much the same way as they adapt to the rate of sentences, by comparing the duration of that action to the rate of surrounding information. Thus, viewers adapt to the speed of various actors when deciding the final action that they saw. This also opens up additional research pathways, such as probing, say, perceptual learning of the rate that particular actors might perform actions at. One could imagine that the speed of one person’s double tap is not the speed of the next. Someone with arthritis, for example, might find quick movements painful, and a viewer’s knowledge of the consequences of this could shape their perception of the actions of someone with arthritis. Could viewers learn those patterns?

3.3.2.2 Relevance to Domain-Specificity

This line of research has clear relevance to questions of domain-specificity. An effect observed in the speech domain was replicated in the domain of event perception. But what was shared between the domains, if anything? The answer could still be "nothing". After all, observing an effect in one domain and a similar effect in another domain does not prove that the two effects are, underlyingly, one and the same. Whale sharks and whales may look similar to each other, eat similar things, and live in similar locations, but one is a fish and the other is a mammal. Similarly, the evolution of eyes in vertebrates and in cephalopods followed similar lines, leading to eyes that are remarkably similar in structure to each other. Although these similarities reveal something interesting about optics and about evolution, it does not imply that, say, the shape of the eyes is programmed by identical genes or identical ontogenetic mechanisms. Still, it may be informative to speculate about how rate adaptation in event perception might be related to rate adaptation in speech to consider how domain-specificity might be studied.

One way in which event perception and speech perception might be linked would be through *representational* structure. That is, there are fundamental commonalities between the representations of speech information and event information. It is unlikely, although possible, that such commonalities would reflect *shared* structure; that is, where events and words are somehow represented in an identical level for some type of processing. However, it is possible to conceive of a *parallel* structure between speech perception and event perception. Although events at different time scales can be referred to as "fine-grained" or "coarse-grained" (Tversky & Zacks, 2013; Zacks & Swallow, 2007), they are all referred to as "events" within the event perception literature. But this need not necessarily be the case. For example, there might be a hierarchy of event structure that resembles that found in the representation of prosody, where segments make up syllables that make up feet that make up words, and so on (Shattuck-Hufnagel & Turk, 1996). It might be helpful to develop a new vocabulary within event segmentation to describe analogous structural properties. Here, speech provides an excellent source of predictions about the structure and form of these representations. This single study alone does not provide nearly enough information to determine the likelihood of representational parallels, however.

Where this study can provide more-direct evidence about domain-specificity relates to the *processes* involved in understanding speech and in understanding events. As discussed in Chapter 1, this is one of the primary determinants of whether a process is considered "modular" in the sense of massive modularity. Does the process of perceiving rate in events resemble that of perceiving rate in speech, as shown through the perception of different events in precursor-rate-modified action sequences? The answer to that seems to be "yes", at least within the bounds of the single experiment performed here. The duration of an event is perceived relative to its context, just as the duration of a word (or of a single segment) is perceived related to its context. One could event imagine a domain-general timing mechanism that could, in some ways, approach the status of a "module". The perception of time is something that has clear evolutionary importance, and managing multiple different "timekeepers" for each independent function would be challenging indeed.

Of course, analogous or similar processing does not automatically imply identical processing; to get closer to that conclusion would require additional experiments. Within event segmentation proper, such experiments could involve manipulating rate or other cues to event boundaries or event identity in a way analogous to the perception of rate. Indeed, one could even imagine cross-modal rate adaptation experiments. If the processes underlying rate adaptation in events and in speech are really one and the same, it should be possible to change the perception of events by manipulating the rate of speech played before the events. Such a possibility is alluring, and it accords well with, say, the influence of the rate of non-linguistic tones on the perception of segment identity, as studied by Holt and colleagues (Holt, 2005; Wade & Holt, 2005). Still, it should also be noted that degraded speech and tonal analogues of speech do very little to affect the segmentation of ambiguous speech, suggesting that the effects may be more limited for segmentation (Pitt, Szostak, & Dilley, 2016).

Another possible link between timing in the speech domain and timing in the event domain relates to a concept key to both domains: prediction. Turk and Shattuck-Hufnagel (2014) proposed that timing variation in speech largely exists in service of maintaining uniform levels of predictability across time. Talkers speed up when producing predictable speech information, and slow down when producing surprising information. This resembles in many ways approaches to modeling syntactic variation in speech production, where speakers attempt to maintain a uniform density for information by using syntactic constructions that modulate the rate of information being produced (Jaeger, 2010), with more informative portions of sentences being presented at a slower effective rate. Predictability also comes into play in the perception of events. For example, not only do viewers keep track of recurring patterns of events in the world around them, they often perceive boundaries in locations where these recurring expectations are violated (Reynolds et al., 2007). As such, it seems plausible to posit that rate adaptation effects in both domains are undergirded by a shared reliance on predictability when it comes to processing events. Future experiments in both domains would be needed to probe these ideas. For all of these eventualities, however, finding the preliminary evidence for domain-general adaptation abilities that was uncovered here provides a useful pathway to exploring each of these ideas in detail.

4 Review and Conclusions

Before moving on to the implications of the studies in this dissertation, it may be helpful to review the ground already covered. This dissertation began with a review of domain-specificity, focusing especially on the “speech-is-special” hypothesis, the idea that the mechanisms that are used to perceive speech are unique to speech processing alone. I chose to examine this hypotheses with regard to two areas: category learning and rate adaptation. Chapter 2.1 reviewed the literature on category learning inside and outside of language. Chapter 2.2 described a study that assessed phonetic category learning, where evidence for a bias against disjunctive phonetic categories was uncovered. In Chapter 2.3, I looked at an analogous category learning scenario with regard to musical instrument categories and found that the anti-disjunctivity bias was not present for at least one set of non-speech stimuli, which I took to be evidence of speech-specificity in the types of categories that were more easily learnable (if not necessarily the mechanisms that were responsible for learning). For rate adaptation, Chapter 3.1 covered rate adaptation in speech perception, and introduced event perception (in particular, event segmentation) as one area in which rate adaptation might also play a role. Chapter 3.2 described an experiment looking at rate adaptation effects in Arabic. And, finally, Chapter 3.3 examined rate adaptation in event perception, finding evidence for rate adaptation effects on the perception of events sharing many of the same properties of those uncovered for speech. Although small methodological differences as well as the underexplored nature of the cues relevant to event segmentation prevent me from concluding with certainty that the processes responsible for adaptation in event perception are *identical* to those observed in speech, the experiment described in Chapter

3.3 at least provides a useful preliminary sketch of how the two domains might be bridged to make future examination of the idea possible.

4.1 Domain-Specificity

What relationship to the present experiments have with notions of domain-specificity? It may be important to consider the ways in which domain-specificity can be uncovered. Domains may be distinguished from one another in terms of their representations; that is, whether the information in each domain is encoded and stored separately. Or they may be distinguished from one another in terms of the processes that operate on those representations; that is, whether the underlying representations are manipulated in the same way and subject to the same operations, even if the precise inputs and outputs are quite different from one another. Or both might be true; representations and processing streams can, but need not necessarily, pattern together.

For instance, under general auditory theories of phonetic processing, speech sounds are represented in the same way as any other auditory object (Lotto, 2000). Because they are represented identically, it is challenging to build in ways for later processing streams to process speech sounds any differently from other auditory objects, save from top-down expectations coming from one’s knowledge of speech. However, it is perfectly possible to imagine disparately-represented information processed using common mechanisms; although ovens take different raw ingredients and turn them into a variety of possible foods, the heating process is shared between all of them. Similarly, a concept like “addition” operates across several possible sets, even if the representations being acted upon are quite different (e.g., imaginary versus real numbers). It may be the case that different types of auditory categories are stored in different ways, but are processed using identical mechanisms. The present experiments both speak more to the processing side of the domain-specificity equation than the representational side, as the focus in the present experiment was on comparing learning and adaptation across experiments rather than, say, looking at cross-modal effects within a single experiment (or even within a single trial).

The experiments assessing category learning could tell a few stories related to the domain-specificity of learning. From the experiments described in Chapters 2.2 and 2.3, it appears that listeners do not come to the task of learning non-speech categories with the same biases as they come to the task of learning speech categories. Disjunctive categories are harder to learn in the speech perception domain than in non-speech auditory categorization. The dataset is entirely consistent with the idea that different processes are used to categorize speech sounds and non-speech sounds. Perhaps non-speech categories are processed in an exemplar-only way, while speech sound categories are learned using rules. But the dataset is also consistent with the idea that identical processes might underpin category learning in both domains. Learners may be storing exemplars of each category being learned and comparing new instances of each category to previous exemplars (although this seems unlikely, given the discussion at the end of Chapter 2.2), positing abstract categories represented by simple rules, or storing prototypical category members in exactly the same way across the domains. Despite identical learning *processes,* it may just be the case that the (identical) processes are influenced by the differing representations of speech and non-speech sounds.

This can most clearly be illustrated through dual-system models of category learning. If the composition of each system is different between speech sounds and non-speech sounds, or even if the composition is identical but the dimensions that are used to represent the categories are different between the domains, it might be that the rule-based system is used more in one domain and the similarity-based system is used more in the other, or that the similarity-based system takes less time to begin processing the sounds in one domain than the other. Simply put, there are many ways that identical processing mechanisms can lead to different results between the two domains if those mechanisms have multiple states. Of course, one of the clearest ways in which this could be true relates to the difference in expertise between speech sounds and musical instrument sounds. Even English speakers likely had at least some passive exposure to the palatal and velar fricative categories, but almost certainly had none with spectrally-rotated musical instrument sounds. Yet English and German speakers did not strongly differ from one another in their acquisition of German fricative categories suggests that expertise is not responsible for the differences between phonetic and musical instrument sound categories. As such, assessing the provenance of the processes used to learn categories will require additional study.

For rate adaptation, meanwhile, it is clear that the main contribution of the present experiments relates more to the processes at play for rate adaptation. Despite their similarities on some levels which I described in some detail in Chapter 3.1, it seems fairly absurd to imagine events and (prosodic) words represented in the same way such that they would both be affected by rate adaptation. Although it is interesting to imagine that the two domains would have *parallel* representations, that idea does not imply *identical* representations. Instead, it seems far more likely that common processes undergird event perception and speech perception. Something about events is processed in the same way as something about speech that this commonality leads to rate adaptation effects.

What could be shared, then, in that processing stream? One idea would be a domain-general conception of time. Time is of the utmost importance in understanding behavior; after all, without a conception of time, the idea of "contingency" becomes meaningless (Gallistel & King, 2010). One could imagine that the perception of rate in events and speech is subject to a single, common timekeeper that tracks the *relative* rate of instances in the world (i.e., tracks whether tempo is speeding up or slowing down), rather than *absolute* rate (i.e., the raw tempo). What instances would be considered "relevant" by this timekeeper to its rate tracking could be different from domain to domain. For example, a unit like the syllable might be a relevant unit of timing for speech, while an analogously relevant unit for event perception would be an interesting object of follow-up studies. Regardless of the relevant unit, however, the timekeeper would be computing the ratio of, say, single syllables to the running average of syllable durations.

Given this idea, though, it is still somewhat challenging to explain the effects of the rate of non-linguistic pure tones on speech perception (Wade & Holt, 2005) without recourse to shared representations, as it is challenging to see why non-linguistic tones should be considered relevant points of comparison for phonetic sequences. In the case of the non-linguistic tone experiments, the *only* information in the immediate context of the syllable is the stream of tones. Perhaps the non-linguistic tones are coerced into representations that are then treated as relevant for the purpose of comparing the (inherently relative) length of the syllable. Something similar seems to occur within speech; the rate of speech spoken even by a clearly different talker (coming from a different location, with a different gender) can influence later-occurring speech (R. S. Newman & Sawusch, 2009).

An alternative proposal is to focus on the role of prediction in both speech perception and event perception. In both domains, boundaries are associated with local minima in terms of predictability (Reynolds et al., 2007; J. R. Saffran, Newport, & Aslin, 1996). However, it is more challenging under this proposal to imagine what a domain-general "prediction" process would entail. It is also unclear why both segment and event identification, not just segmentation, would be affected by rate adaptation.

With all that in mind, it may be useful to end this section with more-general comments about domain-specificity. Common processing streams in either or both domains are an interesting and appealing concept. But even some amount of shared processing would not necessarily sideline any possibility of domain-specificity in either domain. The preponderance of evidence from previous studies seems to put to rest of the idea of speech perception being an encapsulated Fodor-module. To put it another way, the mechanisms of speech perception are not preprogrammed as if they resulted from an instruction manual, describing in detail exactly how different components fit together and leaving no room for adjustment. But I think that it is premature to conclude just because there is no rigid instruction manual that domain-specific attributes are missing entirely.

My view is to find more of a middle ground. I see domain-specific attributes of perceptual processing as a scaffold that other learning is built upon. For instance, biases in speech perception (such as the one here proposed to militate against non-disjunctive categories) could be remarkably *useful* for the learner if they help cut down on improbable hypotheses about speech sounds. The hypothesis space of speech sound categories, for instance, is limited by the functions of the speech apparatus. Indeed, such biases could also be useful in other domains where learning needs to be equivalently quick or where there are similar constraints on possible category members. Conversely, it might also be interesting to study the acquisition of discontinuous categories in sign language phonology, where the theoretical space of possible productions has different boundaries. The idea of domain-specific attributes serving as a scaffold could better fuel further research on the topic of domain-specificity.

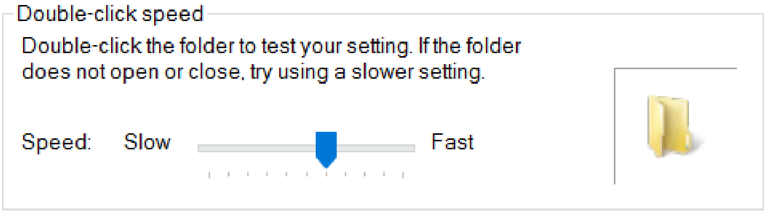
4.2 Applications

Besides its application to the theoretical issues considered above, the research discussed as a part of the present dissertation also lends itself to applications. Below, I consider possible relevance to pedagogical, technological, and clinical applications.

Learning categories—either linguistic or non-linguistic—requires practice. Natural categories are generally not as simple as those acquired in the lab; even ones as complex at the Picket Fence and Odd One Out conditions used in the present experiments involve a unidimensional category structure. The type of category learning theory that one subscribes to can lead to different predictions about the best way to teach natural categories. Under a single-system exemplar-only model of category learning, where category learning is a process strongly predicated on learning individual category instances, it makes sense that teaching people to differentiate individual items can help them acquire broader categories. An example of this comes from the categorization of rock types. Although rocks differ along many visual dimensions—color, grain size, banding, and so on—geologists classify them into one of three types (igneous, metamorphic, and sedimentary) based on the properties of their formation. This categorization is often taught in earth science classrooms. Indeed, when giving individuals instruction about these higher-level categories of rocks, having learners simultaneously acquire both the labels of the higher-level categories and subtype-level labels (gneiss, breccia, obsidian, etc.) improved participants' later categorization to a level greater than that inspired by the higher-level labels alone (Nosofsky, Sanders, Gerdom, Douglas, & McDaniel, 2017).

The results in Chapter 2.2 suggest that exemplar-only accounts of category learning differ from the observed patterns of learning in that they predict that all of the categories being taught could (given enough exposure, and given successful enough discriminability between items) be learned. This mismatch suggests that other learning pathways might be more fruitful for teaching speech-sound categories. Dual-system accounts of category learning predict that learners should shift between different learning systems over the course of learning, with rule-based learning predominating at the beginning of learning and a later transition to similarity-based learning; perhaps those properties could be brought to bear in second language acquisition. Instruction about learning novel speech sounds in a new language could begin with clear, rule-based descriptions of the sound categories in question, reinforced by detailed feedback and plenty of time to process the new information (attributes said to benefit the rule-based system), while later stages of training would involve a switch to quick, simple feedback that is more useful for the similarity-based system.

Meanwhile, the technological implications of the rate adaptation experiments might also be particularly interesting. Most operating systems that I know of have a tool that looks something like the one depicted in Figure 19 buried somewhere in the specifications for the mouse input, where the user can set the duration between two successive clicks where the click is counted as a single action. This is a crude sort of rate normalization. The user has to specify the rate at which two actions must be produced at to be considered one. This means that each user has to specify at a global level what rate the interface should expect successive taps or clicks would occur at without any ability to build in responsiveness to variability across time or across users in the rate of the actions being executed.


Double-click speed

Double-click the folder to test your setting. If the folder does not open or close, try using a slower setting.

Speed: Slow Fast

Figure 19. A double-click sensitivity setting

But what if this process were automated? What if a computer could adapt to the rate that someone clicked at, typed at, or moved a mouse at? Could that improve accessibility? For example, young children sometimes struggle to perform complex mouse actions, such as clicking and dragging (Agudo, Sánchez, & Rico, 2010). Clearly, this is an application that relates more to the production of motor gestures than the perception of them, but one can also imagine rate adaptation also being important for automated motion sensing interfaces, such as Microsoft’s Kinect. If someone producing gestures is seen to be relatively slow (or relatively dynamic), and this fact is used to predict later actions on the part of the actor, this could improve the motion sensor’s ability to pick up on the actions being performed. The parallels between Chapters 3.2

and 3.3 suggest that these processes, and the methods of adapting to them, could possibly be shared between the speech and visual domains.

A final application of this research is clinical. One of the advantages of dual-system theories of category learning is their rich connections to neuropsychology. If phonetic categories are acquired using much the same mechanisms as categories outside the speech realm, dual-system theories predict that the same neuropsychological deficits that lead to impaired non-speech category learning should also apply to phonetic category learning.

According to dual-system theories, the basal ganglia are of particular importance to category learning (Lim et al., 2014). One interesting and underexplored area of research, then, relates to the phonetic category learning abilities (and, indeed, the phonetic perception abilities more generally) of people with Parkinson's disease, a disorder characterized by significant deterioration in the substantia nigra within the basal ganglia. Although the language *production* of people with Parkinson's has been a relatively frequent target of research (Illes, 1989; Illes, Metter, Hanson, & Iritani, 1988), *perception* is another matter. To the extent that dual-system theories can inform phonetic category learning, then, people with Parkinson's disease may show deficits in effective phonetic category learning. This could make it harder for people with Parkinson's disease to learn new speech sound categories; perhaps relatively rare with regard to the acquisition of entirely new *languages*, but more common with regard to contact with speakers of an unfamiliar dialect, or non-native speakers of a first language who have differences from native speakers in the categories being used. Of course, this is predicated on the idea that category learning and adaptation share common mechanisms, an idea highlighted in greater detail below.

There is another population of interest with significant basal ganglia disturbances that is more typically associated with language: people who stutter. People who stutter often differ from controls in the activity and functional connectivity of the basal ganglia during language production (Alm, 2004; Giraud et al., 2008; Watkins, Smith, Davis, & Howell, 2008). Stuttering, then, should correlate with differences in both linguistic and non-linguistic category learning that result from deficits in the dopaminergic system. In one study that is being conducted in collaboration with Soo-Eun Chang at the University of Michigan and Liz Wieland of Michigan State University, I am comparing children who stutter to children who do not stutter in their category learning abilities. So far, children who stutter find rule-based categories particularly challenging to learn when compared to controls. Interestingly, stuttering has also been associated with disturbances in the perception of rhythm, closely linked to the perception of rate (Wieland, McAuley, Dilley, & Chang, 2015), perhaps because the basal ganglia are also important for the perception of timing and rate information (Harrington, Haaland, & Hermanowicz, 1998). Could people who stutter also find rate adaptation more challenging than people who do not? And what does it mean for rate perception to be co-localized with category learning more generally? The implications of this are explored further below.

4.3 Future Directions

Besides applied dimensions, the research discussed in this dissertation also can be further advanced in terms of basic research. I summarize some of the possible continuations below.

4.3.1 Category Learning

In category learning, my research described in Chapters 2.2 and 2.3 suggests that there are differences between linguistic and non-linguistic categories in the types of categories that are easily learnable. Disjunctive categories were more difficult to learn than their non-disjunctive counterparts in the fricative continuum, while disjunctive categories were no more difficult than their non-disjunctive counterparts in the musical instrument continuum. But what are the limits of this bias? Even German speakers, who have extensive experience with the back fricatives, did not show strikingly different patterns from English speakers when learning categories within the [ç]-[x] continuum. Does the anti-disjunctivity bias, then, reflect something unique to those fricatives, or would it be true for other speech sound categories? Would English speakers show similar patterns even for speech sounds completely unlike those used phonetically in English, such as clicks? Or, alternatively, is it musical instruments that are the odd categories out, rather than fricatives? It is readily possible to think of disjunctive categories in music; notes of the scale are disjunctive insofar as a “C” is a “C” whether it is played at 261.6 Hz, 130.8 Hz, or 4186 Hz. Examining other non-linguistic sound categories, such as environmental sounds (keys jangling, animal sounds, etc.) could help examine this possibility. Indeed, these experiments would naturally lend themselves to the sorts of cross-species experimentation that are frequently performed by proponents of general auditory theories of speech perception. If songbirds show similar constraints on the acquisition of disjunctive categories to human, this suggests that it is likely not something unique to human representations of speech (or experience with speech) that leads to the bias against disjunctive categories within the fricative continuum.

A second future path of study would be to probe the predictions of different theories as to how people might learn these categories. As discussed extensively in Chapter 2.2, exemplar-only views of category learning are not easily capable of explaining the anti-disjunctivity bias. However, I hedged for a reason: it *may* be possible to do so given a relatively complex system of attentional focus and complex dimensionality within the stimuli. Still, other category learning models may do a better job of accommodating the bias. But which model to choose? Certainly, dual-system models are promising; besides being the object of recent study in the field of speech perception (Maddox & Chandrasekaran, 2014), they also come with a prepackaged roster of diagnostic tests said to distinguish between the rule-based and similarity-based learning systems (Maddox & Ashby, 2004). Yet dual-system models in the visual domain sometimes make the claim that rules in the rule-based system can be disjunctive in nature (Minda et al., 2008), which would sharply undercut my proposal that the difference between the two systems explains the difference between the disjunctive and non-disjunctive categories in the present experiment. Careful experimentation could tease these different possibilities apart by, for example, determining perceived dimensionality before and after the acquisition of each type of phonetic category.

Careful study of the dimensions of phonetic space could also aid in circling back to some of the parallel theoretical questions not covered in detail in this paper, such as the nature of phonetic representations (i.e., auditory or motor). I make no claims about whether the categories being learned within the fricative continuum are disjunctive in terms of *motor* space or in terms of *auditory* space because, at least theoretically, the continuum used here confounded the auditory information in the fricatives with the motor gestures that would have led to the intermediate fricative steps. One could imagine using tokens that are disjunctive in auditory space but non-disjunctive in motor space (or vice-versa) to study which of those categories are more difficult than the others. The present experiments also only explored a putatively unidimensional category learning continuum. Phonetic categories are multidimensional in nature Examining categories that are, say, disjunctive on one dimension and non-disjunctive on another could provide an intriguing path to assess the nature of these phonetic dimensions, as well as to push the boundaries of this anti-disjunctive bias.

4.3.2 Rate Adaptation

The opportunities are in some ways even more unbounded in segmentation. Just the apparent use of rate adaptation for event perception alone provides a wide variety of possible future explorations. The rate adaptation examined in the present experiment involved manipulating the rate of every action but the critical one. Could rate adaptation be demonstrated using only a subset of the context; say, for example, an analogue of the distal context, as examined in Chapter 3.2? Could listeners show perceptual learning, picking up the idea that a certain action tends to be executed more slowly for an actor? Could adaptation happen across actors, events, or even modalities? Data that could start answering these questions could be provided using the current stimulus set alone, to say nothing about the studying these phenomena using a broader and more naturalistic set of actions besides just ones executed on a touchscreen.

However, limiting exploration of the parallels of word segmentation and event segmentation to context rate is unwise. Connections between event segmentation and word segmentation have only recently been put into the literature (Peña et al., 2011), and there are numerous parallels between event perception and speech perception that remain to be explored. For instance, the interplay between top-down and bottom-up cues is one that has been tentatively explored in event segmentation (Bridgette M Hard et al., 2006; Zacks, 2004). One influential review of the literature on event segmentation still described "the detailed characterization of the relation between bottom-up and top-down processing in event segmentation as one important goal for future research" (Zacks & Swallow, 2007, p. 83). This relationship is one that has been extensively explored in the word segmentation literature. One might ask if the same sort of hierarchy of segmentation cues similar to that proposed by Mattys et al. (2005) for word segmentation might also be applicable to event segmentation, with top-down cues taking priority over bottom-up cues. This awaits investigation of each of these classes of cues, both separately and in combination.

4.3.3 Synthesis

Throughout this dissertation, I have discussed category learning and rate adaptation in parallel to one another: parallel literature reviews, parallel previous experiments, and parallel applications of the experiments within the domain of language to problems outside of language. I hope that I have succeeded in providing insight into questions of domain-specificity in both fields. However, I believe there is more to the connection between learning and adaptation than just parallels. After all, the human brain is plastic: it changes with experience, which allows behavior to vary in line with changing situational demands. Both learning new sound categories in non-native languages (here referred to as "phonetic learning") and adjusting native phonetic categories to accommodate differences from speakers of other dialects or accents (here, “phonetic adaptation”) require brains to be plastic. Could it be that both phonetic learning and phonetic adaptation are supported by common mechanisms of plasticity?

The idea of a common framework for phonetic learning and phonetic adaptation is driven by two motivating factors. First, a parsimonious approach to model building suggests that language learners will use the existing architectures available from their native language to learn new speech sounds. Thus, adaptation, traditionally situated within the first language capacity of the learner, could be used in order to acquire categories in the second language. Second, there are hints that some aspects of phonetic learning and phonetic adaptation are influenced by similar factors. For instance, sleep consolidation significantly improves both phonetic learning (Earle & Myers, 2015) and dialect adaptation (Fenn, Margoliash, & Nusbaum, 2013; Fenn, Nusbaum, & Margoliash, 2003), which suggests that sleep consolidation may work on similar neural substrates for each task. Unifying phonetic learning and phonetic adaptation could lead to a richer understanding of both phenomena, allowing for new connections between the theoretical tools used to explore each phenomenon independently as well as new hypotheses related to the neural architecture of phonetic plasticity. Similar approaches have already been undertaken with regard to the representation of action sequences (Botvinick & Plaut, 2004) and syntactic structures (Chang, Dell, & Bock, 2006; Dell & Chang, 2014; Jaeger & Snider, 2013).

Under this notion, the same abilities that allow listeners to track statistical distributions in one’s native language also allow listeners to posit distributions for categories in other languages. This best resembles proposals by Toscano and others (Toscano & McMurray, 2010) that emphasize the importance of cue reliability in both the acquisition and maintenance of phonetic categories. Dual-system models of phonetic learning may be able to provide that neural description for both phonetic learning (Chandrasekaran, Koslov, et al., 2014; Chandrasekaran, Yi, et al., 2014; Maddox & Chandrasekaran, 2014) *and* phonetic adaptation. Several underexplored predictions fall out of this model. For example, one idea is that the variation in the properties of the rule-based learning system (and corresponding frontal areas) does not contribute to meaningful variation in individual speech plasticity abilities. Another idea is that subcortical areas may exert meaningful influence on phonetic adaptation in addition to phonetic learning (Lim et al., 2014). Such an idea, though, would take time, and a great deal of new research, to demonstrate.

4.4 Closing Thoughts

The present dissertation took an interdisciplinary approach to categorization and adaptation inside and outside speech perception. Certainly, the virtues of an approach that applies principles from cognitive science theories to language science has become readily apparent to language scientists; witness, for example, Pierrehumbert's (2003) wholesale adoption of Nosofsky's (1986) theories of categorization into the speech literature. These are efforts to be lauded. But I also believe that language scientists equally have something to offer cognitive science more generally, even beyond simply an interesting set of topics to study, as important as that is. Rather, language scientists have struggled with many of the important topics of cognitive science for years; issues such as categorization and adaptation come up again and again in language. And this amassed knowledge base can be used to generate hypotheses relevant to many other cognitive domains. It is time for language scientists to stop being bashful about this fact, and to start broadcasting their knowledge to the broader scientific community. To the extent that the present dissertation helps amplify this broadcast, I have succeeded in my goals.

5 References

Adank, P., Evans, B. G., Stuart-Smith, J., & Scott, S. K. (2009). Comprehension of familiar and unfamiliar native accents under adverse listening conditions. *Journal of Experimental Psychology: Human Perception and Performance, 35*(2), 520–529. http://doi.org/10.1037/a0013552

Agudo, J. E., Sánchez, H., & Rico, M. (2010). Playing games on the screen: Adapting mouse interaction at early ages. In M. Jemni Kinshuk, D. Sampson, & J. M. Spector (Eds.), *Proceedings of the 10th IEEE International Conference on Advanced Learning Technologies* (pp. 493–497). Sousse, Tunisia: IEEE. http://doi.org/10.1109/ICALT.2010.142

Al-Tamimi, F. Y. (2004). An experimental phonetic study of intervocalic singleton and geminate sonorants in Jordanian Arabic. *Al-'Arabiyya*, *37*, 37–52.

Al-Tamimi, J., & Khattab, G. (2011). Multiple cues for the singleton-geminate contrast in Lebanese Arabic: Acoustic investigation of stops and fricatives. In *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 212–215). Hong Kong.

Alm, P. A. (2004). Stuttering and the basal ganglia circuits: A critical review of possible relations. *Journal of Communication Disorders*, *37*(4), 325–369. http://doi.org/10.1016/j.jcomdis.2004.03.001

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review*, *98*(3), 409–429.

Anderson, J. R., & Betz, J. (2001). A hybrid model of categorization. *Psychonomic Bulletin & Review*, *8*(4), 629–647. http://doi.org/10.3758/BF03196200

Andrews, T. J., Schluppeck, D., Homfray, D., Matthews, P., & Blakemore, C. (2002).

Activity in the fusiform gyrus predicts conscious perception of Rubin's vase-face illusion. *NeuroImage*, *17*(2), 890–901. http://doi.org/10.1016/S1053-8119(02)91243-7

Ashby, F. G., & Alfonso-Reese, L. A. (1995). Categorization as probability density estimation. *Journal of Mathematical Psychology, 39*(2), 216–233.

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, *105*(3), 442–481. http://doi.org/10.1037/0033-295X.105.3.442

Ashby, F. G., & Ell, S. W. (2002). Single versus multiple systems of category learning: Reply to Nosofsky and Kruschke (2002). *Psychonomic Bulletin & Review*, *9*(1), 175–180. http://doi.org/10.3758/BF03196275

Ashby, F. G., & Gott, R. E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(1), 33–53. http://doi.org/10.1037/0278-7393.14.1.33

Ashby, F. G., & Maddox, W. T. (1993). Relations between prototype, exemplar, and decision bound models of categorization. *Journal of Mathematical Psychology*, *37*(3), 372–400. http://doi.org/10.1006/jmps.1993.1023

Ashby, F. G., & Maddox, W. T. (2005). Human category learning. *Annual Review of Psychology*, *56*, 149–178. http://doi.org/10.1146/annurev.psych.56.091103.070217

Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review*, *93*(2), 154–179.

Ashby, F. G., & Waldron, E. M. (1999). On the nature of implicit categorization. *Psychonomic Bulletin & Review*, *6*(3), 363–378. http://doi.org/10.3758/BF03210826

Ashby, F. G., Waldron, E. M., Lee, W. W., & Berkman, A. (2001). Suboptimality in human categorization and identification. *Journal of Experimental Psychology: General*, *130*(1), 77–96. http://doi.org/10.1037/0096-3445.130.1.77

Auerbach, S. H., Allard, T., Naeser, M., Alexander, M. P., & Albert, M. L. (1982). Pure word deafness: Analysis of a case with bilateral lesions and a defect at the prephonemic level. *Brain*, *105*(2), 271–300. http://doi.org/10.1093/brain/105.2.271

Baayen, R. H., Bates, D. M., Kliegl, R., & Vasishth, S. (2015). RePsychLing: Data sets from Psychology and Linguistics experiments. Retrieved from https://github.com/dmbates/RePsychLing

Baese-Berk, M. M., Bradlow, A. R., & Wright, B. A. (2013). Accent-independent adaptation to foreign accented speech. *Journal of the Acoustical Society of America*, *133*(3), EL174-EL180.

Baese-Berk, M. M., Heffner, C. C., Dilley, L. C., Pitt, M. A., Morrill, T. H., & McAuley, J. D. (2014). Long-term temporal tracking of speech rate affects spoken-word recognition. *Psychological Science*, *25*(8), 1546–1553. http://doi.org/10.1177/0956797614533705

Bahrick, L. E., Gogate, L. J., & Ruiz, I. (2002). Attention and memory for faces and actions in infancy: The salience of actions over faces in dynamic events. *Child Development*, *73*(6), 1629–1643. http://doi.org/10.1111/1467-8624.00495

Baldwin, D. A., Andersson, A., Saffran, J., & Meyer, M. (2008). Segmenting dynamic human action via statistical structure. *Cognition*, *106*(3), 1382–1407. http://doi.org/10.1016/j.cognition.2007.07.005

Baldwin, D. A., Baird, J. A., Saylor, M. M., & Clark, M. A. (2001). Infants parse dynamic action. *Child Development*, *72*(3), 708–717. http://doi.org/10.1111/1467-8624.00310

Bates, D. M., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. *ArXiv*.

Bates, D. M., Maechler, M., Bolker, B., & Walker, S. (2016). lme4: Linear mixed-effects models using Eigen and S4. Retrieved from http://cran.r-project.org/package=lme4

Batterink, L., & Neville, H. (2011). Implicit and explicit mechanisms of word learning in a narrative context: An event-related potential study. *Journal of Cognitive Neuroscience*, *23*(11), 3181–3196. http://doi.org/10.1162/jocn_a_00013

Beach, E. F., Burnham, D., & Kitamura, C. (2001). Bilingualism and the relationship between perception and production: Greek/English bilinguals and Thai bilabial stops. *International Journal of Bilingualism*, *5*(2), 221–235.

Benson, R. R., Whalen, D. H., Richardson, M., Swainson, B., Clark, V. P., Lai, S., & Liberman, A. M. (2001). Parametrically dissociating speech and nonspeech perception in the brain using fMRI. *Brain and Language*, *78*(3), 364–396. http://doi.org/10.1006/brln.2001.2484

Bent, T., & Bradlow, A. R. (2003). The interlanguage speech intelligibility benefit. *Journal of the Acoustical Society of America*, *114*(3), 1600–1610. http://doi.org/10.1121/1.1603234

Berent, I., Pinker, S., & Shimron, J. (2002). The nature of regularity and irregularity: Evidence from Hebrew nominal inflection. *Journal of Psycholinguistic Research*, *31*(5), 459–502. http://doi.org/10.1023/A:1021256819323

Best, C. T. (1995). A direct realist view of cross-language speech perception. In W.

Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171–206). Timonium, MD: York Press.

Best, C. T., Tyler, M. D., Gooding, T. N., Orlando, C. B., & Quann, C. A. (2009). Development of phonological constancy. *Psychological Science*, *20*(5), 539–542. http://doi.org/10.1111/j.1467-9280.2009.02327.x

Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S., Springer, J. A., Kaufman, J. N., & Possing, E. T. (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cerebral Cortex*, *10*(5), 512–528. http://doi.org/10.1093/cercor/10.5.512

Bion, R. A. H., Benavides-Varela, S., & Nespor, M. (2011). Acoustic markers of prominence influence infants' and adults' segmentation of speech sequences. *Language and Speech*, *54*(1), 123–140. http://doi.org/10.1177/0023830910388018

Blumstein, S. E., & Stevens, K. N. (1979). Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants. *Journal of the Acoustical Society of America*, *66*(4), 1001–1017. http://doi.org/10.1121/1.383319

Boersma, P., & Weenink, D. (2001). Praat: Doing phonetics by computer. *Glot International*, *5*(9/10), 341–345.

Boggia, J., & Ristic, J. (2015). Social event segmentation. *Quarterly Journal of Experimental Psychology*, *68*(4), 731–744. http://doi.org/10.1080/17470218.2014.964738

Bohnemeyer, J., Enfield, N. J., Essegbey, J., Ibarretxe-Antuñano, I., Kita, S., Lüpke, F., & Ameka, F. K. (2007). Principles of event segmentation in language: The case of motion events. *Language*, *83*(3), 495–532. http://doi.org/10.1353/lan.2007.0116

Bortfeld, H., Morgan, J. L., Golinkoff, R. M., & Rathbun, K. (2005). Mommy and me. *Psychological Science*, *16*(4), 298–304.

Botvinick, M. M., & Plaut, D. C. (2004). Doing without schema hierarchies: A recurrent connectionist approach to normal and impaired routine sequential action. *Psychological Review*, *111*(2), 395–429. http://doi.org/10.1037/0033-295X.111.2.395

Bowden, H. W., Gelfand, M. P., Sanz, C., & Ullman, M. T. (2010). Verbal inflectional morphology in L1 and L2 Spanish: A frequency effects study examining storage versus composition. *Language Learning*, *60*(1), 44–87. http://doi.org/10.1111/j.1467-9922.2009.00551.x

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, *106*(2), 707–729. http://doi.org/10.1016/j.cognition.2007.04.005

Bruderer, A. G., Danielson, D. K., Kandhadai, P., & Werker, J. F. (2015). Sensorimotor influences on speech perception in infancy. *Proceedings of the National Academy of Sciences*, *112*(44), 13531–13536. http://doi.org/10.1073/pnas.1508631112

Bukach, C. M., Gauthier, I., & Tarr, M. J. (2006). Beyond faces and modularity: the power of an expertise framework. *Trends in Cognitive Sciences*, *10*(4), 159–166. http://doi.org/10.1016/j.tics.2006.02.004

Burdick, C. K., & Miller, J. D. (1975). Speech perception by the chinchilla: discrimination of sustained /a/ and /i/. *Journal of the Acoustical Society of America*, *58*(2), 415–427. http://doi.org/10.1121/1.381006

Busemeyer, J. R., Dewey, G. I., & Medin, D. L. (1984). Evaluation of exemplar-based generalization and the abstraction of categorical information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(4), 638–648. http://doi.org/10.1037//0278-7393.7.6.418

Buxó-Lugo, A., & Watson, D. G. (2016). Evidence for the influence of syntax on prosodic parsing. *Journal of Memory and Language*, *90*, 1–13. http://doi.org/10.1016/j.jml.2016.03.001

Bybee, J. L. (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change*, *14*(3), 261–290. http://doi.org/10.1017/S0954394502143018

Carroll, J. D., & Chang, J.-J. (1970). Analysis of individual differences in multidimensional scaling via an n-way generalization of “Eckart-Young” decomposition. *Psychometrika*, *35*(3), 283–319.

Carruthers, P. (2005). The Case for massively modular models of mind. In R. J. Stainton (Ed.), *Contemporary Debates in Cognitive Science* (pp. 3–21). Oxford: Blackwell. http://doi.org/10.1111/j.1468-0017.2008.00340.x

Chandrasekaran, B., Koslov, S. R., & Maddox, W. T. (2014). Toward a dual-learning systems model of speech category learning. *Frontiers in Psychology*, *5*, 825. http://doi.org/10.3389/fpsyg.2014.00825

Chandrasekaran, B., Sampath, P. D., & Wong, P. C. M. (2010). Individual variability in cue-weighting and lexical tone learning. *Journal of the Acoustical Society of America*, *128*(1), 456–465. http://doi.org/10.1121/1.3445785

Chandrasekaran, B., Yi, H.-G., Blanco, N. J., McGeary, J. E., & Maddox, W. T. (2015). Enhanced procedural learning of speech sound categories in a genetic variant of

FOXP2. *Journal of Neuroscience*, *35*(20), 7808–7812. http://doi.org/10.1523/JNEUROSCI.4706-14.2015

Chandrasekaran, B., Yi, H.-G., & Maddox, W. T. (2014). Dual-learning systems during speech category learning. *Psychonomic Bulletin & Review*, *21*(2), 488–495. http://doi.org/10.3758/s13423-013-0501-5

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review*, *113*(2), 234–272. http://doi.org/10.1037/0033-295X.113.2.234

Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *Journal of the Acoustical Society of America*, *116*(6), 3647–3658. http://doi.org/10.1121/1.1815131

Cohen Kadosh, K., & Johnson, M. H. (2007). Developing a cortex specialized for face perception. *Trends in Cognitive Sciences*, *11*(9), 367–369. http://doi.org/10.1016/j.tics.2007.06.007

Colombo, J., & Bundy, R. S. (1981). A method for the measurement of infant auditory selectivity. *Infant Behavior and Development*, *4*(1), 219–223. http://doi.org/10.1016/S0163-6383(81)80025-2

Conti-Ramsden, G., Ullman, M. T., & Lum, J. A. G. (2015). The relation between receptive grammar and procedural, declarative, and working memory in specific language impairment. *Frontiers in Psychology*, *6*, 1090. http://doi.org/10.3389/fpsyg.2015.01090

Cosmides, L., & Tooby, J. (1994). Origins of domain specificity: The evolution of functional organization. In L. Hirschfeld & S. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 85–111). New York: Cambridge

University Press.

Craik, F. I. M., & Hay, J. F. (1999). Aging and judgments of duration: Effects of task complexity and method of estimation. *Perception & Psychophysics*, *61*(3), 549–560. http://doi.org/10.3758/BF03211972

Cristia, A., Seidl, A., Vaughn, C., Schmale, R., Bradlow, A., & Floccia, C. (2012). Linguistic processing of accented speech across the lifespan. *Frontiers in Psychology*, *3*, 479. http://doi.org/10.3389/fpsyg.2012.00479

Crystal, T. H., & House, A. S. (1988). Segmental durations in connected-speech signals: Current results. *Journal of the Acoustical Society of America*, *83*(4), 1553–1573. http://doi.org/10.1121/1.395911

Curtin, S., Mintz, T. H., & Christiansen, M. H. (2005). Stress changes the representational landscape: Evidence from word segmentation. *Cognition*, *96*(3), 233–262. http://doi.org/10.1016/j.cognition.2004.08.005

Cutting, J. E. (2014). Event segmentation and seven types of narrative discontinuity in popular movies. *Acta Psychologica*, *149*, 69–77. http://doi.org/10.1016/j.actpsy.2014.03.003

Cutting, J. E., Brunick, K. L., & Candan, A. (2012). Perceiving event dynamics and parsing Hollywood films. *Journal of Experimental Psychology: Human Perception and Performance, 38*(6), 1476–1490. http://doi.org/10.1037/a0027737

Davis, M. H., Di Betta, A. M., Macdonald, M. J. E., & Gaskell, M. G. (2009). Learning and consolidation of novel spoken words. *Journal of Cognitive Neuroscience, 21*(4), 803–820. http://doi.org/10.1162/jocn.2009.21059

Davis, M. H., & Gaskell, M. G. (2009). A complementary systems account of word learning: Neural and behavioural evidence. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *364*(1536), 3773–3800. http://doi.org/10.1098/rstb.2009.0111

Davis, M. H., Marslen-Wilson, W. D., & Gaskell, M. G. (2002). Leading up the lexical garden path: Segmentation and ambiguity in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(1), 218–244. http://doi.org/10.1037//0096-1523.28.1.218

De Renzi, E. (1986). Prosopagnosia in two patients with CT scan evidence of damage confined to the right hemisphere. *Neuropsychologia*, *24*(3), 385–389. http://doi.org/10.1016/0028-3932(86)90023-0

Deen, B., Richardson, H., Dilks, D. D., Takahashi, A., Keil, B., Wald, L. L., … Saxe, R. R. (2017). Organization of high-level visual cortex in human infants. *Nature Communications*, *8*, 13995. http://doi.org/10.1038/ncomms13995

Delattre, P. C., Liberman, A. M., & Cooper, F. S. (1955). Acoustic loci and transitional cues for consonants. *Journal of the Acoustical Society of America*, *27*(4), 769–773.

Dell, G. S., & Chang, F. (2014). The P-chain: Relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *369*(1634), 20120394. http://doi.org/10.1098/rstb.2012.0394

Dent, M. L., Brittan-Powell, E. F., Dooling, R. J., & Pierce, A. (1997). Perception of synthetic /ba/-/wa/ speech continuum by budgerigars (Melopsittacus undulatus). *Journal of the Acoustical Society of America*, *102*(3), 1891–1897. http://doi.org/10.1121/1.420111

Diehl, R. L., Lotto, A. J., & Holt, L. L. (2004). Speech perception. *Annual Review of Psychology*, *55*, 149–179. http://doi.org/10.1146/annurev.psych.55.090902.142028

Diehl, R. L., Souther, A. F., & Convis, C. L. (1980). Conditions on rate normalization in speech perception. *Perception & Psychophysics*, *27*(5), 435–443. http://doi.org/10.3758/BF03204461

Dilley, L. C., Mattys, S. L., & Vinke, L. (2010). Potent prosody: Comparing the effects of distal prosody, proximal prosody, and semantic context on word segmentation. *Journal of Memory and Language*, *63*(3), 274–294. http://doi.org/10.1016/j.jml.2010.06.003

Dilley, L. C., & McAuley, J. D. (2008). Distal prosodic context affects word segmentation and lexical processing. *Journal of Memory and Language*, *59*(3), 294–311. http://doi.org/10.1016/j.jml.2008.06.006

Dilley, L. C., Morrill, T. H., & Banzina, E. (2013). New tests of the distal speech rate effect: Examining cross-linguistic generalization. *Frontiers in Psychology, 4*, 1002. http://doi.org/10.3389/fpsyg.2013.01002

Dilley, L. C., & Pitt, M. A. (2010). Altering context speech rate can cause words to appear or disappear. *Psychological Science*, *21*(11), 1664–1670. http://doi.org/10.1177/0956797610384743

Donkin, C., Newell, B. R., Kalish, M., Dunn, J. C., & Nosofsky, R. M. (2015). Identifying strategy use in category learning tasks: A case for more diagnostic data and models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(4), 933–948. http://doi.org/10.1037/xlm0000083

Dooling, R. J., Best, C. T., & Brown, S. D. (1995). Discrimination of synthetic full-formant and sinewave /ra-la/ continua by budgerigars (Melopsittacus undulatus) and zebra finches (Taeniopygia guttata). *Journal of the Acoustical Society of America*, *97*(3), 1839–1846. http://doi.org/10.1121/1.412058

Dooling, R. J., Okanoya, K., & Brown, S. D. (1989). Speech perception by budgerigars (Melopsittacus undulatus): the voiced-voiceless distinction. *Perception & Psychophysics*, *46*(1), 65–71. http://doi.org/10.3758/BF03208075

Dronkers, N. F., Wilkins, D. P., Van Valin, R. D., Redfern, B. B., & Jaeger, J. J. (2004). Lesion analysis of the brain areas involved in language comprehension. *Cognition*, *92*(1–2), 145–177. http://doi.org/10.1016/j.cognition.2003.11.002

Duchaine, B. C., Dingle, K., Butterworth, E., & Nakayama, K. (2004). Normal greeble learning in a severe case of developmental prosopagnosia. *Neuron*, *43*(4), 469–473. http://doi.org/10.1016/j.neuron.2004.08.006

Dumay, N., & Gaskell, M. G. (2007). Sleep-associated changes in the mental representation of spoken words. *Psychological Science*, *18*(1), 35–39. http://doi.org/10.1111/j.1467-9280.2007.01845.x

Dumay, N., & Gaskell, M. G. (2012). Overnight lexical consolidation revealed by speech segmentation. *Cognition*, *123*(1), 119–132. http://doi.org/10.1016/j.cognition.2011.12.009

Dupoux, E., & Green, K. (1997). Perceptual adjustment to highly compressed speech: Effects of talker and rate changes. *Journal of Experimental Psychology: Human Perception and Performance*, *23*(3), 914–927. http://doi.org/10.1037/0096-1523.23.3.914

Earle, F. S., & Myers, E. B. (2015). Sleep and native language interference affect non-native speech sound learning. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(6), 1680–1695.

Egnor, S. E. R., & Hauser, M. D. (2004). A paradox in the evolution of primate vocal learning. *Trends in Neurosciences*, *27*(11), 649–654. http://doi.org/10.1016/j.tins.2004.08.009

Eimas, P. D., & Miller, J. L. (1992). Organization in the perception of speech by young infants. *Psychological Science*, *3*(6), 340–345.

Eisner, F., & McQueen, J. M. (2005). The specificity of perceptual learning in speech processing. *Perception & Psychophysics*, *67*(2), 224–238. http://doi.org/10.3758/BF03206487

Enard, W., Fisher, S. E., Przeworski, M., Lai, C. S. L., Wiebe, V., Kitano, T., … Pääbo, S. (2002). Molecular evolution of FOXP2, a gene involved in speech and language. *Nature, 418*(6900), 869–872. http://doi.org/10.1038/nature01025

Endress, A. D., & Wood, J. N. (2011). From movements to actions: Two mechanisms for learning action sequences. *Cognitive Psychology*, *63*(3), 141–171. http://doi.org/10.1016/j.cogpsych.2011.07.001

Erickson, M. A., & Kruschke, J. K. (1998). Rules and exemplars in category learning. *Journal of Experimental Psychology: General*, *127*(2), 107–140. http://doi.org/10.1037/0096-3445.127.2.107

Erickson, M. A., & Kruschke, J. K. (2002). Rule-based extrapolation in perceptual categorization. *Psychonomic Bulletin & Review, 9*(1), 160–168. http://doi.org/10.3758/BF03196273

Esposito, A., & Di Benedetto, M. G. (1999). Acoustical and perceptual study of gemination in Italian stops. *Journal of the Acoustical Society of America*, *106*(4), 2051–2062. http://doi.org/10.1121/1.428056

Ezzyat, Y., & Davachi, L. (2011). What constitutes an episode in episodic memory? *Psychological Science*, *22*(2), 243–252. http://doi.org/10.1177/0956797610393742

Faber, M., & Gennari, S. P. (2015). In search of lost time: Reconstructing the unfolding of events from memory. *Cognition*, *143*, 193–202. http://doi.org/10.1016/j.cognition.2015.06.014

Faber, M., & Gennari, S. P. (2017). Effects of learned episodic event structure on prospective duration judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. http://doi.org/10.1037/xlm0000378

Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is “special” about face perception? *Psychological Review*, *105*(3), 482–498. http://doi.org/10.1037/0033-295X.105.3.482

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. http://doi.org/10.1037/a0017196.The

Fenn, K. M., Margoliash, D., & Nusbaum, H. C. (2013). Sleep restores loss of generalized but not rote learning of synthetic speech. *Cognition*, *128*(3), 280–286. http://doi.org/10.1016/j.cognition.2013.04.007

Fenn, K. M., Nusbaum, H. C., & Margoliash, D. (2003). Consolidation during sleep of perceptual learning of spoken language. *Nature*, *425*(6958), 614–616. http://doi.org/10.1038/nature01951

Fitch, W. T., de Boer, B., Mathur, N., & Ghazanfar, A. A. (2016). Monkey vocal tracts are speech-ready. *Science Advances, 2,* e1600723. http://doi.org/10.1126/sciadv.1600723

Floccia, C., Butler, J., Goslin, J., & Ellis, L. (2009). Regional and foreign accent processing in English: Can listeners adapt? *Journal of Psycholinguistic Research, 38*(4), 379–412. http://doi.org/10.1007/s10936-008-9097-8

Floccia, C., Goslin, J., Girard, F., & Konopczynski, G. (2006). Does a regional accent perturb speech processing? *Journal of Experimental Psychology: Human Perception and Performance, 32*(5), 1276–1293. http://doi.org/10.1037/0096-1523.32.5.1276

Fodor, J. A. (1983). *The Modularity of Mind: An Essay on Faculty Psychology*. Cambridge, Mass.: MIT Press.

Forrest, K., Weismer, G., Milenkovic, P., & Dougall, R. N. (1988). Statistical analysis of word-initial voiceless obstruents: Preliminary data. *Journal of the Acoustical Society of America*, *84*(1), 115–123. http://doi.org/10.1121/1.396977

Fowler, C. (1990). Sound-producing sources as objects of perception: Rate normalization and nonspeech perception. *Journal of the Acoustical Society of America*, *88*(3), 1236–1249. http://doi.org/10.1121/1.399701

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*, *14*(1), 3–28. Retrieved from http://129.237.66.221/P800/Fowler1986.pdf

Fowler, C. A., & Rosenblum, L. D. (1990). Duplex Perception: A Comparison of Monosyllables and Slamming Doors. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(4), 742–754. http://doi.org/10.1037/0096-1523.16.4.742

Frankenhuis, W. E., & Ploeger, A. (2007). Evolutionary psychology versus Fodor: Arguments for and against the massive modularity hypothesis. *Philosophical Psychology*, *20*(6), 687–710. http://doi.org/10.1080/09515080701665904

Fujii, T., Fukatsu, R., Watabe, S., Ohnuma, A., Teramura, K., Kimura, I., … Kogure, K. (1990). Auditory sound agnosia without aphasia following a right temporal lobe lesion. *Cortex*, *26*(2), 263–268. http://doi.org/10.1001/archneur.1965.00470010088012

Gabay, Y., & Holt, L. L. (2015). Incidental learning of sound categories is impaired in developmental dyslexia. *Cortex*, *73*, 131–143. http://doi.org/10.1016/j.cortex.2015.08.008

Gadian, D. G., Aicardi, J., Watkins, K. E., Porter, D. A., Mishkin, M., & Vargha-Khadem, F. (2000). Developmental amnesia associated with early hypoxic-ischaemic injury. *Brain*, *123*(3), 499–507. http://doi.org/10.1093/brain/123.3.499

Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review*, *13*(3), 361–377. http://doi.org/10.3758/BF03193990

Gallistel, C. R., & King, A. P. (2010). *Memory and the Computational Brain: Why Cognitive Science Will Transform Neuroscience*. *Cognitive Science* (Vol. 3). Chichester, England: Wiley-Blackwell. http://doi.org/10.1002/9781444310498

Gardiner, J. M., Brandt, K. R., Baddeley, A. D., Vargha-Khadem, F., & Mishkin, M. (2008). Charting the acquisition of semantic knowledge in a case of developmental amnesia. *Neuropsychologia*, *46*(11), 2865–2868. http://doi.org/10.1016/j.neuropsychologia.2008.05.021

Gaskell, M. G., & Dumay, N. (2003). Lexical competition and the acquisition of novel words. *Cognition*, *89*(2), 105–132. http://doi.org/10.1016/S0010-0277(03)00070-2

Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience*, *3*(2), 191–197. http://doi.org/10.1038/72140

Gauthier, I., & Tarr, M. J. (1997). Becoming a "greeble" expert: Exploring mechanisms for face recognition. *Vision Research, 37*(12), 1673–1682.

Giraud, A.-L., Neumann, K., Bachoud-Levi, A.-C., von Gudenberg, A. W., Euler, H. A., Lanfermann, H., & Preibisch, C. (2008). Severity of dysfluency correlates with basal ganglia activity in persistent developmental stuttering. *Brain and Language*, *104*, 190–199. http://doi.org/10.1016/j.bandl.2007.04.005

Gold, D. A., Zacks, J. M., & Flores, S. (2017). Effects of cues to event segmentation on subsequent memory. *Cognitive Research: Principles and Implications*, *2*, 1. http://doi.org/10.1186/s41235-016-0043-2

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*(2), 251–279. http://doi.org/10.1037/0033-295X.105.2.251

Goldinger, S. D., & Azuma, T. (2003). Puzzle-solving science: The quixotic quest for units in speech perception. *Journal of Phonetics, 31*(3–4), 305–320. http://doi.org/10.1016/S0095-4470(03)00030-5

Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology*, *49,* 585–612. http://doi.org/10.1146/annurev.psych.49.1.585

Goodman, N. D., Tenenbaum, J. B., Feldman, J., & Griffiths, T. L. (2008). A rational analysis of rule-based concept learning. *Cognitive Science*, *32*(1), 108–154. http://doi.org/10.1080/03640210701802071

Gordon-Salant, S., & Fitzgibbons, P. J. (2001). Sources of age-related recognition difficulty for time-compressed speech. *Journal of Speech, Language, and Hearing Research*, *44*(4), 709–719. http://doi.org/10.1044/1092-4388(2001/056)

Goren, C. C., Sarty, M., & Wu, P. Y. K. (1975). Visual following and pattern discrimination of face-like stimuli by newborn infants. *Pediatrics*, *56*(4), 544–549.

Goslin, J., Duffy, H., & Floccia, C. (2012). An ERP investigation of regional and foreign accent processing. *Brain and Language*, *122*(2), 92–102. http://doi.org/10.1016/j.bandl.2012.04.017

Graf Estes, K., Evans, J. L., Alibali, M. W., & Saffran, J. R. (2007). Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychological Science*, *18*(3), 254–260. http://doi.org/10.1111/j.1467-9280.2007.01885.x

Grieser, D., & Kuhl, P. K. (1989). Categorization of speech by infants: Support for speech-sound prototypes. *Developmental Psychology*, *25*(4), 577–588. http://doi.org/10.1037/0012-1649.25.4.577

Grill-Spector, K., Knouf, N., & Kanwisher, N. (2004). The fusiform face area subserves face perception, not generic within-category identification. *Nature Neuroscience*, *7*(5), 555–562. http://doi.org/10.1038/nn1224

Grosvald, M. (2009). Interspeaker variation in the extent and perception of long-distance vowel-to-vowel coarticulation. *Journal of Phonetics*, *37*(2), 173–188. http://doi.org/10.1016/j.wocn.2009.01.002

Guenther, F. H., Espy-Wilson, C. Y., Boyce, S. E., Matthies, M. L., Zandipour, M., & Perkell, J. S. (1999). Articulatory tradeoffs reduce acoustic variability during American English /r/ production. *Journal of the Acoustical Society of America*, *105*(5), 2854–2865. http://doi.org/10.1121/1.426900

Hankamer, J., Lahiri, A., & Koreman, J. (1989). Perception of consonant length: Voiceless stops in Turkish and Bengali. *Journal Of Phonetics*, *17*, 283–298.

Hard, B. M., Recchia, G., & Tversky, B. (2011). The shape of action. *Journal of Experimental Psychology: General*, *140*(4), 586–604. http://doi.org/10.1037/a0024310

Hard, B. M., Tversky, B., & Lang, D. S. (2006). Making sense of abstract events: Building event schemas. *Memory & Cognition*, *34*(6), 1221–1235. http://doi.org/10.3758/BF03193267

Harrington, D. L., Haaland, K. Y., & Hermanowicz, N. (1998). Temporal processing in the basal ganglia. *Neuropsychology*, *12*(1), 3–12.

Hawkins, S. (2003). Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of Phonetics*, *31*(3–4), 373–405. http://doi.org/10.1016/j.wocn.2003.09.006

Hawkins, S. (2010). Phonological features, auditory objects, and illusions. *Journal of Phonetics*, *38*(1), 60–89. http://doi.org/10.1016/j.wocn.2009.02.001

Hedenius, M., Persson, J., Alm, P. A., Ullman, M. T., Howard, J. H., Howard, D. V, & Jennische, M. (2013). Impaired implicit sequence learning in children with developmental dyslexia. *Research in Developmental Disabilities*, *34*(11), 3924–3935. http://doi.org/10.1016/j.ridd.2013.08.014

Heffner, C. C., Dilley, L. C., McAuley, J. D., & Pitt, M. A. (2013). When cues combine: How distal and proximal acoustic cues are integrated in word segmentation. *Language and Cognitive Processes*, *28*(9), 1275–1302. http://doi.org/10.1080/01690965.2012.672229

Heffner, C. C., Newman, R. S., Dilley, L. C., & Idsardi, W. J. (2015). Age-related differences in speech rate perception do not necessarily entail age-related differences in speech rate use. *Journal of Speech, Language, and Hearing Research*, *58*(4), 1341–1349. http://doi.org/10.1044/2015

Heffner, C. C., Newman, R. S., & Idsardi, W. J. (2017). Support for context effects on segmentation and segments depends on the context. *Attention, Perception, & Psychophysics*.

Hemeren, P. E., & Thill, S. (2011). Deriving motor primitives through action segmentation. *Frontiers in Psychology*, *1*, 243. http://doi.org/10.3389/fpsyg.2010.00243

Henderson, L. M., Weighall, A. R., Brown, H., & Gaskell, M. G. (2012). Consolidation of vocabulary is associated with sleep in children. *Developmental Science*, *15*(5), 674–687. http://doi.org/10.1111/j.1467-7687.2012.01172.x

Hespos, S. J., Grossman, S. R., & Saylor, M. M. (2010). Infants' ability to parse continuous actions: Further evidence. *Neural Networks*, *23*(8–9), 1026–1032. http://doi.org/10.1016/j.neunet.2010.07.010

Hespos, S. J., Saylor, M. M., & Grossman, S. R. (2009). Infants' ability to parse continuous actions. *Developmental Psychology*, *45*(2), 575–585. http://doi.org/10.1016/j.neunet.2010.07.010

Hickok, G. (2009). Eight problems for the mirror neuron theory of action understanding in monkeys and humans. *Journal of Cognitive Neuroscience*, *21*(7), 1229–1243. http://doi.org/10.1162/jocn.2009.21189

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews: Neuroscience*, *8*(5), 393–402. http://doi.org/10.1038/nrn2113

Hintzman, D. L. (1986). “Schema abstraction” in a multiple-trace memory model. *Psychological Review*. http://doi.org/10.1037/0033-295X.93.4.411

Hirschfeld, L. A., & Gelman, S. A. (1994). Toward a topography of mind: An introduction to domain specificity. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the Mind: Domain Specificity in Cognition and Culture* (pp. 3–13). New York: Cambridge University Press.

Hirsh-Pasek, K., Kemler Nelson, D. G., Jusczyk, P. W., Cassidy, K. W., Druss, B., & Kennedy, L. (1987). Clauses are perceptual units for young infants. *Cognition*, *26*(3), 269–286. http://doi.org/10.1016/S0010-0277(87)80002-1

Hodgson, P., & Miller, J. L. (1996). Internal structure of phonetic categories: Evidence for within-category trading relations. *Journal of the Acoustical Society of America*, *100*(1), 565–576. http://doi.org/10.1121/1.415867

Hohne, E. A., & Jusczyk, P. W. (1994). Two-month-old infants’ sensitivity to allophonic differences. *Perception & Psychophysics*, *56*(6), 613–623. http://doi.org/10.3758/BF03208355

Holt, L. L. (2005). Temporally nonadjacent nonlinguistic sounds affect speech categorization. *Psychological Science*, *16*(4), 305–312. http://doi.org/10.1111/j.0956-7976.2005.01532.x

Holt, L. L. (2006). The mean matters: effects of statistically defined nonspeech spectral distributions on speech categorization. *Journal of the Acoustical Society of America*, *120*(5), 2801–2817. http://doi.org/10.1121/1.2354071

Holt, L. L., & Kluender, K. R. (2000). General auditory processes contribute to perceptual accommodation of coarticulation. *Phonetica*, *57*, 170–180. http://doi.org/10.1159/000028470

Holt, L. L., & Lotto, A. J. (2008). Speech perception wthin an auditory cognitive science framework. *Current Directions in Psychological Science*, *17*(1), 42–46. http://doi.org/10.1111/j.1467-8721.2008.00545.x

Holt, L. L., & Lotto, A. J. (2010). Speech perception as categorization. *Attention, Perception, & Psychophysics*, *72*(5), 1218–1227. http://doi.org/10.3758/APP.72.5.1218

Holt, L. L., Lotto, A. J., & Diehl, R. L. (2004). Auditory discontinuities interact with categorization: Implications for speech perception. *Journal of the Acoustical Society of America*, *116*(3), 1763–1773. http://doi.org/10.1121/1.1778838

Holt, L. L., Lotto, A. J., & Kluender, K. R. (2000). Neighboring spectral content influences vowel identification. *Journal of the Acoustical Society of America*, *108*(2), 710–722. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10955638

Holt, L. L., Lotto, A. J., & Kluender, K. R. (2001). Influence of fundamental frequency on stop-consonant voicing perception: A case of learned covariation or auditory enhancement? *Journal of the Acoustical Society of America*, *109*(2), 764–774. http://doi.org/10.1121/1.427666

Homa, D., Cross, J., Cornell, D., Goldman, D., & Shwartz, S. (1973). Prototype abstraction and classification of new instances as a function of number of instances defining the prototype. *Journal of Experimental Psychology*, *101*(1), 116–122. http://doi.org/10.1037/h0035772

Homa, D., Sterling, S., & Trepel, L. (1981). Limitations of exemplar-based generalization and the abstraction of categorical information. *Journal of Experimental Psychology: Human Learning and Memory*, *7*(6), 418–439. http://doi.org/10.1037//0278-7393.7.6.418

Idemaru, K., & Guion-Anderson, S. (2010). Relational timing in the production and perception of Japanese singleton and geminate stops. *Phonetica*, *67*(1–2), 25–46. http://doi.org/10.1159/000319377

Idemaru, K., & Guion, S. G. (2008). Acoustic covariants of length contrast. *Journal of the International Phonetic Association*, *38*(2), 167–186. http://doi.org/10.1017/S0025100308003459

Idemaru, K., Holt, L. L., & Seltman, H. (2012). Individual differences in cue weights are stable across time: The case of Japanese stop lengths. *Journal of the Acoustical Society of America*, *132*(6), 3950–3964.

Illes, J. (1989). Neurolinguistic features of spontaneous language production dissociate three forms of neurodegenerative disease: Alzheimer’s, Huntington’s, and Parkinson’s. *Brain and Language*. http://doi.org/10.1016/0093-934X(89)90116-8

Illes, J., Metter, E. J., Hanson, W. R., & Iritani, S. (1988). Language production in Parkinson’s disease: Acoustic and linguistic considerations. *Brain and Language, 33*(1), 146–160. http://doi.org/10.1016/0093-934X(88)90059-4

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. http://doi.org/10.1121/1.412280

Iverson, P., Wagner, A., & Rosen, S. (2016). Effects of language experience on pre-categorical perception: Distinguishing general from specialized processes in speech perception. *Journal of the Acoustical Society of America*, *139*(4), 1799–1809. http://doi.org/10.1121/1.4944755

Jacobs, R. A., & Kruschke, J. K. (2011). Bayesian learning theory applied to human cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(1), 8–21. http://doi.org/10.1002/wcs.80

Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, *61*(1), 23–62. http://doi.org/10.1016/j.cogpsych.2010.02.002

Jaeger, T. F., & Snider, N. E. (2013). Alignment as a consequence of expectation adaptation: Syntactic priming is affected by the prime’s prediction error given both prior and recent experience. *Cognition*, *127*(1), 57–83. http://doi.org/10.1016/j.cognition.2012.10.013

Janse, E., & Adank, P. (2012). Predicting foreign-accent adaptation in older adults. *Quarterly Journal of Experimental Psychology*, *65*(8), 1563–1585.

Jarvis, E. D. (2004). Learned birdsong and the neurobiology of human language. *Annals of the New York Academy of Sciences*, *1016*, 749–777. http://doi.org/10.1196/annals.1298.038

Johnson, E. K., & Jusczyk, P. W. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language*, *44*, 548–

567. http://doi.org/10.1006/jmla.2000.2755

Jusczyk, P. W., & Aslin, R. N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology*, *29*(1), 1–23. http://doi.org/10.1006/cogp.1995.1010

Jusczyk, P. W., Hirsh-Pasek, K., Kemler Nelson, D. G., Kennedy, L. J., Woodward, A., & Piwoz, J. (1992). Perception of acoustic correlates of major phrasal units by young infants. *Cognitive Psychology*, *24*(2), 252–293. http://doi.org/10.1016/0010-0285(92)90009-Q

Jusczyk, P. W., Hohne, E. A., & Bauman, A. (1999). Infants' sensitivity to allophonic cues for word segmentation. *Perception & Psychophysics*, *61*(8), 1465–1476. http://doi.org/10.3758/BF03213111

Jusczyk, P. W., Smolensky, P., & Allocco, T. (2002). How English-learning infants respond to markedness and faithfulness constraints. *Language Acquisition*, *10*(1), 31–73. http://doi.org/10.1207/S15327817LA1001

Kalish, M. L., & Kruschke, J. K. (1997). Decision boundaries in one-dimensional categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(6), 1362–1377. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9372605

Kanwisher, N. (2000). Domain specificity in face perception. *Nature Neuroscience*, *3*(8), 759–763. http://doi.org/10.1038/77664

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, *17*(11), 4302–4311. http://doi.org/10.1098/Rstb.2006.1934

Kapnoula, E. C., & McMurray, B. (2016). Newly learned word forms are abstract and integrated immediately after acquisition. *Psychonomic Bulletin & Review*, *23*(2), 491–499. http://doi.org/10.3758/s13423-015-0897-1

Kapnoula, E. C., Packard, S., Gupta, P., & McMurray, B. (2015). Immediate lexical integration of novel word forms. *Cognition*, *134*, 85–99. http://doi.org/10.1016/j.cognition.2014.09.007

Kazanina, N., Phillips, C., & Idsardi, W. (2006). The influence of meaning on the perception of speech sounds. *Proceedings of the National Academy of Sciences*, *103*(16), 11381–11386. http://doi.org/10.1073/pnas.0604821103

Kemp, C., Perfors, A., & Tenenbaum, J. B. (2007). Learning overhypotheses with hierarchical Bayesian models. *Developmental Science*, *10*(3), 307–321. http://doi.org/10.1111/j.1467-7687.2007.00585.x

Keren, G., & Schul, Y. (2009). Two is not always better than one: A critical evaluation of two-system theories. *Perspectives on Psychological Science*, *4*(6), 533–550.

Kessinger, R. H., & Blumstein, S. E. (1998). Effects of speaking rate on voice-onset time and vowel production: Some implications for perception studies. *Journal of Phonetics*, *26*(2), 117–128. http://doi.org/10.1006/jpho.1997.0069

Key, M. (2014). Positive expectation in the processing of allophones. *Journal of the Acoustical Society of America*, *135*(6), EL350-EL356. http://doi.org/10.1121/1.4879669

Kidd, E. (2012). Implicit statistical learning is directly associated with the acquisition of syntax. *Developmental Psychology*, *48*(1), 171–184. http://doi.org/10.1037/a0025405

Kidd, G. R. (1989). Articulatory-rate context effects in phoneme identification. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(4), 736–748. http://doi.org/10.1037/0096-1523.15.4.736

Kingston, J., Kawahara, S., Chambless, D., Mash, D., & Brenner-Alsop, E. (2009). Contextual effects on the perception of duration. *Journal of Phonetics, 37*(3), 297–320. http://doi.org/10.1016/j.wocn.2009.03.007

Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review, 122*(2), 148–203. http://doi.org/10.1037/a0038695

Klin, A., Jones, W., Schultz, R., Volkmar, F., & Cohen, D. (2002). Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Archives of General Psychiatry, 59*(9), 809–816. http://doi.org/10.1001/archpsyc.59.9.809

Kraljic, T., Brennan, S. E., & Samuel, A. G. (2008). Accommodating variation: Dialects, idiolects, and speech processing. *Cognition*, *107*(1), 54–81. http://doi.org/10.1016/j.cognition.2007.07.013

Kraljic, T., & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology*, *51*(2), 141–178. http://doi.org/10.1016/j.cogpsych.2005.05.001

Kraljic, T., & Samuel, A. G. (2006). Generalization in perceptual learning for speech. *Psychonomic Bulletin & Review*, *13*(2), 262–268. http://doi.org/10.3758/BF03193841

Kronrod, Y., Coppess, E., & Feldman, N. H. (2016). A unified model of categorical effects in phonetic perception. *Psychonomic Bulletin & Review*, *23*(6), 1681–1712. http://doi.org/10.3758/s13423-016-1049-y

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, *99*(1), 22–44. http://doi.org/10.1037/0033-295X.99.1.22

Kruschke, J. K. (2005). Category learning. In K. Lamberts & R. L. Goldstone (Eds.), *The Handbook of Cognition* (pp. 183–201). London: SAGE. http://doi.org/10.1016/B978-0-12-397025-1.00274-8

Kruschke, J. K. (2008). Models of categorization. In R. Sun (Ed.), *The Cambridge Handbook of Computational Psychology* (pp. 267–301). New York: Cambridge University 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 & Psychophysics*, *50*(2), 93–107. http://doi.org/10.3758/BF03212211

Kuhl, P. K., & Miller, J. D. (1975). Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants. *Science*, *190*(4209), 69–72. http://doi.org/10.1126/science.1166301

Kuhl, P. K., & Miller, J. D. (1978). Speech perception by the chinchilla: Identification function for synthetic VOT stimuli. *Journal of the Acoustical Society of America*, *63*(3), 905–917. http://doi.org/10.1121/1.381770

Kurby, C. A., & Zacks, J. M. (2008). Segmentation in the perception and memory of events. *Trends in Cognitive Sciences*, *12*(2), 72–79. http://doi.org/10.1016/j.tics.2007.11.004

Kurby, C. A., & Zacks, J. M. (2011). Age differences in the perception of hierarchical structure in events. *Memory & Cognition*, *39*(1), 75–91. http://doi.org/10.3758/s13421-010-0027-2

Kurdziel, L. B. F., Mantua, J., & Spencer, R. M. C. (2016). Novel word learning in older adults: A role for sleep? *Brain and Language*, *167*, 106–113. http://doi.org/10.1016/j.bandl.2016.05.010

Lacerda, F. (1995). The perceptual-magnet effect: An emergent consequence of exemplar-based phonetic memory. *Proceedings of the XIIIth International Congress of Phonetic Sciences*, *2*, 140–147. Retrieved from http://www.sfs.uni-tuebingen.de/~gjaeger/lehre/ss08/exemplarBased/Lacerda95.pdf

Ladefoged, P., & Halle, M. (1988). Some major features of the International Phonetic Alphabet. *Language*, *64*(3), 577–582.

Lambert, J., Eustache, F., Lechevalier, B., Rossa, Y., & Viader, F. (1989). Auditory agnosia with relative sparing of speech perception. *Cortex*, *25*(1), 71–82.

Lamberts, K. (2000). Information-accumulation theory of speeded categorization. *Psychological Review*, *107*(2), 227–260. http://doi.org/10.1037//0033-295X.107.2.227

Lane, H. (1965). The motor theory of speech perception. *Psychological Review*, *72*(4), 275–309. http://doi.org/10.1016/0010-0277(85)90021-6

Lassaline, M. E., & Murphy, G. L. (1998). Alignment and category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(1), 144–160. http://doi.org/10.1037/0278-7393.24.1.144

Lavric, A., Pizzagalli, D., Forstmeier, S., & Rippon, G. (2001). Mapping dissociations in verb morphology. *Trends in Cognitive Sciences*, *5*(7), 301–308. http://doi.org/10.1016/S1364-6613(00)01703-4

Leach, L., & Samuel, A. G. (2007). Lexical configuration and lexical engagement: When adults learn new words. *Cognitive Psychology*, *55*(4), 306–353. http://doi.org/10.1016/j.cogpsych.2007.01.001

Leaver, A. M., & Rauschecker, J. P. (2010). Cortical representation of natural complex sounds: Effects of acoustic features and auditory object category. *Journal of Neuroscience*, *30*(22), 7604–7612. http://doi.org/10.1523/JNEUROSCI.0296-10.2010

Lenth, R. (2016). lsmeans: Least-Squares Means. Retrieved from https://cran.r-project.org/package=lsmeans

Levi, S. V. (2015). Individual differences in learning talker categories: The role of working memory. *Phonetica*, *71*, 201–226. http://doi.org/10.1159/000370160

Liberman, A. M. (1957). Some results of research on speech perception. *Journal of the Acoustical Society of America*, *29*(1), 117–123. http://doi.org/10.1121/1.1908635

Liberman, A. M. (1982). On finding that speech is special. *American Psychologist*, *37*(2), 148–167. http://doi.org/10.1037//0003-066X.37.2.148

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*(6), 431–461. http://doi.org/10.1037/h0020279

Liberman, A. M., Delattre, P., & Cooper, F. S. (1952). The rôle of selected stimulus-variables in the perception of the unvoiced stop consonants. *American Journal of Psychology*, *65*(4), 497–516. Retrieved from http://www.jstor.org/stable/1418032?seq=1#page_scan_tab_contents

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*(1), 1–36. http://doi.org/10.3758/PBR.15.2.453

Liberman, A. M., & Mattingly, I. G. (1989). A specialization for speech perception. *Science*, *243*(4890), 489–494. http://doi.org/10.1126/science.2740894

Lim, S.-J., Fiez, J. A., & Holt, L. L. (2014). How may the basal ganglia contribute to auditory categorization and speech perception? *Frontiers in Neuroscience*, *8*, 230. http://doi.org/10.3389/fnins.2014.00230

Lindsay, S., & Gaskell, M. G. (2010). A complementary systems account of word learning in L1 and L2. *Language Learning*, *60*(Suppl. 2), 45–63.

Lindsay, S., & Gaskell, M. G. (2013). Lexical integration of novel words without sleep. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(2), 608–622. http://doi.org/10.1037/a0029243

Liu, R., & Holt, L. L. (2011). Neural changes associated with nonspeech auditory category learning parallel those of speech category acquisition. *Journal of Cognitive Neuroscience*, *23*(3), 683–698. http://doi.org/10.1162/jocn.2009.21392

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.

Livingston, K. R., Andrews, J. K., & Harnad, S. (1998). Categorical perception effects induced by category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(3), 732–753.

Lotto, A. J. (2000). Language acquisition as complex category formation. *Phonetica*, *57*(2–4), 189–196. http://doi.org/28472

Lotto, A. J., Hickok, G. S., & Holt, L. L. (2009). Reflections on mirror neurons and speech perception. *Trends in Cognitive Sciences*, *13*(3), 110–114. http://doi.org/10.1016/j.tics.2008.11.008

Lotto, A. J., & Kluender, K. R. (1998). General contrast effects in speech perception: effect of preceding liquid on stop consonant identification. *Perception & Psychophysics*, *60*(4), 602–619. http://doi.org/10.3758/BF03206049

Lotto, A. J., Kluender, K. R., & Holt, L. L. (1997). Perceptual compensation for coarticulation by Japanese quail (Coturnix coturnix japonica). *Journal of the Acoustical Society of America*, *102*(2), 1134–1140. http://doi.org/10.1121/1.419865

Lu, S., Harter, D., & Graesser, A. C. (2009). An empirical and computational investigation of perceiving and remembering event temporal relations. *Cognitive Science*, *33*(3), 345–373. http://doi.org/10.1111/j.1551-6709.2009.01016.x

Lum, J. A. G., Conti-Ramsden, G., Page, D., & Ullman, M. T. (2012). Working, declarative and procedural memory in specific language impairment. *Cortex*, *48*(9), 1138–1154. http://doi.org/10.1016/j.cortex.2011.06.001

Maddox, W. T., & Ashby, F. G. (2004). Dissociating explicit and procedural-learning based systems of perceptual category learning. *Behavioural Processes*, *66*(3), 309–332. http://doi.org/10.1016/j.beproc.2004.03.011

Maddox, W. T., & Chandrasekaran, B. (2014). Tests of a dual-system model of speech category learning. *Bilingualism: Language and Cognition*, *17*(4), 709–728. http://doi.org/10.1017/S1366728913000783

Maddox, W. T., Chandrasekaran, B., Smayda, K., & Yi, H.-G. (2013). Dual systems of speech category learning across the lifespan. *Psychology and Aging*, *28*(4), 1042–1056. http://doi.org/10.1037/a0034969

Maddox, W. T., Chandrasekaran, B., Smayda, K., Yi, H.-G., Koslov, S., & Beevers, C. G. (2014). Elevated depressive symptoms enhance reflexive but not reflective auditory category learning. *Cortex*, *58*, 186–198. http://doi.org/10.1016/j.cortex.2014.06.013

Maddox, W. T., Filoteo, J. V., Hejl, K. D., & Ing, A. D. (2004). Category number impacts rule-based but not information-integration category learning: Further evidence for dissociable category-learning systems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(1), 227–245. http://doi.org/10.1037/0278-7393.30.1.227

Magliano, J. P., & Zacks, J. M. (2011). The impact of continuity editing in narrative film on event segmentation. *Cognitive Science*, *35*(8), 1489–1517. http://doi.org/10.1111/j.1551-6709.2011.01202.x

Magnuson, J. S., Tanenhaus, M. K., Aslin, R. N., & Dahan, D. (2003). The time course of spoken word learning and recognition: studies with artificial lexicons. *Journal of Experimental Psychology: General*, *132*(2), 202–227. http://doi.org/10.1037/0096-3445.132.2.202

Mair, P., De Leeuw, J., Borg, I., & Groenen, P. J. F. (2016). smacof: Multidimensional Scaling. Retrieved from https://cran.r-project.org/web/packages/smacof/index.html

Mann, V. A. (1986). Distinguishing universal and language-dependent levels of speech perception: Evidence from Japanese listeners' perception of English "l" and "r." *Cognition*, *24*(3), 169–196. http://doi.org/10.1016/S0010-0277(86)80001-4

Mann, V. A., & Liberman, A. M. (1983). Some differences between phonetic and auditory modes of perception. *Cognition*, *14*(2), 211–235.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: W. H. Freeman.

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, *10*(1), 29–63. http://doi.org/10.1016/0010-0285(78)90018-X

Massaro, D. W., & Oden, G. C. (1980). Evaluation and integration of acoustic features in speech perception. *Journal of the Acoustical Society of America*, *67*(3), 996–1013. http://doi.org/10.1121/1.383941

Mattys, S. L. (1997). The use of time during lexical processing and segmentation: A review. *Psychonomic Bulletin & Review*, *4*(3), 310–329. http://doi.org/10.3758/BF03210789

Mattys, S. L., Jusczyk, P. W., Luce, P. A., & Morgan, J. L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology*, *38*(4), 465–494. http://doi.org/10.1006/cogp.1999.0721

Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General*, *134*(4), 477–500. http://doi.org/10.1037/0096-3445.134.4.477

Maye, J., Aslin, R. N., & Tanenhaus, M. K. (2008). The Weckud Wetch of the Wast: Lexical adaptation to a novel accent. *Cognitive Science*, *32*, 543–562. http://doi.org/10.1080/03640210802035357

Maye, J., Weiss, D. J., & Aslin, R. N. (2008). Statistical phonetic learning in infants:

Facilitation and feature generalization. *Developmental Science*, *11*(1), 122–134. http://doi.org/10.1111/j.1467-7687.2007.00653.x

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*(3), 101–111. http://doi.org/10.1016/S0010-0277(01)00157-3

McAuley, J. D., Jones, M. R., Holub, S., Johnston, H. M., & Miller, N. S. (2006). The time of our lives: Life span development of timing and event tracking. *Journal of Experimental Psychology: General*, *135*(3), 348–367. http://doi.org/10.1037/0096-3445.135.3.348

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1–86. http://doi.org/10.1016/0010-0285(86)90015-0

McKinley, S. C., & Nosofsky, R. M. (1995). Investigations of exemplar and decision bound models in large, ill-defined category structures. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(1), 128–148. http://doi.org/10.1037/0096-1523.21.1.128

McKone, E., Kanwisher, N., & Duchaine, B. C. (2007). Can generic expertise explain special processing for faces? *Trends in Cognitive Sciences*, *11*(1), 8–15. http://doi.org/10.1016/j.tics.2006.11.002

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. http://doi.org/10.1111/j.1467-7687.2009.00822.x

McMurray, B., Tanenhaus, M. K., Aslin, R. N., & Spivey, M. J. (2003). Probabilistic constraint satisfaction at the lexical/phonetic interface: Evidence for gradient effects of within-category VOT on lexical access. *Journal of Psycholinguistic Research*, *32*(1), 77–97.

McNeil, J. E., & Warrington, E. K. (1993). Prosopagnosia: A face-specific disorder. *Quarterly Journal of Experimental Psychology*, *46A*(1), 1–10. http://doi.org/10.1080/14640749308401064

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*(3), 207–238. http://doi.org/10.1037/0033-295X.85.3.207

Mervis, C. B., & Rosch, E. H. (1981). Categorization of natural objects. *Annual Review of Psychology*, *32*, 89–115. http://doi.org/10.1146/annurev.ps.32.020181.000513

Meyer, M., Zysset, S., von Cramon, D. Y., & Alter, K. (2005). Distinct fMRI responses to laughter, speech, and sounds along the human peri-sylvian cortex. *Cognitive Brain Research*, *24*(2), 291–306. http://doi.org/10.1016/j.cogbrainres.2005.02.008

Miller, J. D. (1970). Audibility curve of the chinchilla. *Journal of the Acoustical Society of America*, *48*(2), 513–523. http://doi.org/10.1121/1.1914321

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* (pp. 39–74). Hillsdale, NJ.

Miller, J. L. (1994). On the internal structure of phonetic categories: A progress report. *Cognition*, *50*(1–3), 271–285. http://doi.org/10.1016/0010-0277(94)90031-0

Miller, J. L. (1997). Internal structure of phonetic categories. *Language and Cognitive Processes*, *12*(5–6), 865–870. http://doi.org/10.1080/016909697386754

Miller, J. L., Connine, C. M., Schermer, T. M., & Kluender, K. R. (1983). A possible auditory basis for internal structure of phonetic categories. *Journal of the Acoustical Society of America*, *73*(6), 2124–2133. http://doi.org/10.1121/1.389455

Miller, J. L., & Dexter, E. R. (1988). Effects of speaking rate and lexical status on phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(3), 369–378. http://doi.org/10.1037//0096-1523.14.3.369

Miller, J. L., & Eimas, P. D. (1983). Studies on the categorization of speech by infants. *Cognition*, *13*(2), 135–165. http://doi.org/10.1016/0010-0277(83)90020-3

Miller, J. L., & Liberman, A. M. (1979). Some effects of later-occurring information on the perception of stop consonant and semivowel. *Perception & Psychophysics*, *25*(6), 457–465. http://doi.org/10.3758/BF03213823

Miller, J. L., & Volaitis, L. E. (1989). Effect of speaking rate on the perceptual structure of a phonetic category. *Perception & Psychophysics*, *46*(6), 505–512. http://doi.org/10.3758/BF03208147

Minda, J. P., Desroches, A. S., & Church, B. A. (2008). Learning rule-described and non-rule-described categories: A comparison of children and adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(6), 1518–1533. http://doi.org/10.1037/a0013355

Moreton, E., Pater, J., & Pertsova, K. (2017). Phonological concept learning. *Cognitive Science*, *41*(1), 4–69. http://doi.org/10.1111/cogs.12319

Morgan-Short, K., Sanz, C., Ullman, M. T., & Steinhauer, K. (2010). Second Language Acquisition of Gender Agreement in Explicit and Implicit Training Conditions: An Event-Related Potential Study. *Language Learning*, *60*(1), 154–193. http://doi.org/10.1111/j.1467-9922.2009.00554.x.Second

Morgan-Short, K., Steinhauer, K., Sanz, C., & Ullman, M. T. (2012). Explicit and implicit second language training differentially affect the achievement of native-like brain activation patterns. *Journal of Cognitive Neuroscience*, *24*(4), 933–947. http://doi.org/10.1162/jocn_a_00119

Morrill, T. H., Baese-Berk, M. M., Heffner, C. C., & Dilley, L. C. (2015). Interactions between distal speech rate, linguistic knowledge, and speech environment. *Psychonomic Bulletin & Review*, *22*(5), 1451–1457. http://doi.org/10.3758/s13423-015-0820-9

Moscovitch, M., Winocur, G., & Behrmann, M. (1997). What is special about face recognition? Nineteen experiments on a person with visual object agnosia and dyslexia but normal face recognition. *Journal of Cognitive Neuroscience*, *9*(5), 555–604.

Myers, E. B. (2014). Emergence of category-level sensitivities in non-native speech sound learning. *Frontiers in Neuroscience*, *8*, 238. http://doi.org/10.3389/fnins.2014.00238

Myers, J., Jusczyk, P. W., Kemler Nelson, D. G., Charles-Luce, J., Woodward, A. L., & Hirsh-Pasek, K. (1996). Infants' sensitivity to word boundaries in fluent speech. *Journal of Child Language*, *23*(1), 1–30. http://doi.org/10.1017/S0305000900010072

Nakai, S., & Scobbie, J. M. (2016). The VOT category boundary in word-initial stops: Counter-evidence against rate normalization in English spontaneous speech. *Laboratory Phonology*, *7*(1), 13.

Newell, B. R. (2012). Levels of explanation in category learning. *Australian Journal of Psychology*, *64*(1), 46–51. http://doi.org/10.1111/j.1742-9536.2011.00035.x

Newell, B. R., Dunn, J. C., & Kalish, M. (2011). Systems of category learning: Fact or fantasy? In B. Ross (Ed.), *Psychology of Learning and Motivation: Advances in Research and Theory* (Vol. 54, pp. 167–215). Burlington: Academic Press. http://doi.org/10.1016/B978-0-12-385527-5.00006-1

Newman, A. J., Ullman, M. T., Pancheva, R., Waligura, D. L., & Neville, H. J. (2007). An ERP study of regular and irregular English past tense inflection. *NeuroImage*, *34*(1), 435–445. http://doi.org/10.1016/j.neuroimage.2006.09.007

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*(3), 1181–1196. http://doi.org/10.1121/1.1348009

Newman, R. S., Ratner, N. B., Jusczyk, A. M., Jusczyk, P. W., & Dow, K. A. (2006). Infants' early ability to segment the conversational speech signal predicts later language development: A retrospective analysis. *Developmental Psychology*, *42*(4), 643–655. http://doi.org/10.1037/0012-1649.42.4.643

Newman, R. S., & Sawusch, J. R. (1996). Perceptual normalization for speaking rate: Effects of temporal distance. *Perception & Psychophysics*, *58*(4), 540–560. http://doi.org/10.3758/BF03213089

Newman, R. S., & Sawusch, J. R. (2009). Perceptual normalization for speaking rate III: Effects of the rate of one voice on perception of another. *Journal of Phonetics*, *37*(1), 46–65. http://doi.org/10.1016/j.wocn.2008.09.001.Perceptual

Newtson, D. (1973). Attribution and the unit of perception of ongoing behavior. *Journal of Personality and Social Psychology*, *28*(1), 28–38. http://doi.org/10.1037/h0035584

Nielbo, K. L., & Sørensen, J. (2011). Spontaneous processing of functional and nonfunctional action sequences. *Religion, Brain & Behavior, 1*(1), 18–30. http://doi.org/10.1080/2153599X.2010.550722

Norman, K. A., & O'Reilly, R. C. (2003). Modeling hippocampal and neocortical contributions to recognition memory: A complementary-learning-systems approach. *Psychological Review*, *110*(4), 611–646. http://doi.org/10.1037/0033-295X.110.4.611

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, *47*(2), 204–238. http://doi.org/10.1016/S0010-0285(03)00006-9

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, *115*(1), 39–61. http://doi.org/10.1037/0096-3445.115.1.39

Nosofsky, R. M., & Kruschke, J. K. (2002). Single-system models and interference in category learning: Commentary on Waldron and Ashby (2001). *Psychonomic Bulletin & Review*, *9*(1), 169–174. http://doi.org/10.3758/BF03196274

Nosofsky, R. M., Kruschke, J. K., & McKinley, S. C. (1992). Combining exemplar-based category representations and connectionist learning rules. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(2), 211–233.

Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review*, *104*(2), 266–300. http://doi.org/10.1037/0033-295X.104.2.266

Nosofsky, R. M., & Palmeri, T. J. (1998). A rule-plus-exception model for classifying objects in continuous-dimension spaces. *Psychonomic Bulletin & Review*, *5*(3), 345–

369.

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review*, *101*(1), 53–79. http://doi.org/10.1037/0033-295X.101.1.53

Nosofsky, R. M., Sanders, C. A., Gerdom, A., Douglas, B. J., & McDaniel, M. A. (2017). On learning natural-science categories that violate the family resemblance principle. *Psychological Science*, *28*(1), 104–114. http://doi.org/10.1177/0956797616675636

Obrecht, D. H. (1965). Three experiments in the perception of geminate consonants in Arabic. *Language and Speech*, *8*(1), 31–41.

Ongchoco, J. D. K., Uddenberg, S., & Chun, M. M. (2016). Statistical learning of movement. *Psychonomic Bulletin & Review*, *23*(6), 1913–1919. http://doi.org/10.3758/s13423-016-1046-1

Orfanidou, E., Marslen-Wilson, W. D., & Davis, M. H. (2006). Neural response suppression predicts repetition priming of spoken words and pseudowords. *Journal of Cognitive Neuroscience*, *18*(8), 1237–1252. http://doi.org/10.1162/jocn.2006.18.8.1237

Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(2), 309–328.

Palmeri, T. J., Wong, A. C.-N., & Gauthier, I. (2004). Computational approaches to the development of perceptual expertise. *Trends in Cognitive Sciences*, *8*(8), 378–386. http://doi.org/10.1016/j.tics.2004.06.001

Pascalis, O., de Haan, M., & Nelson, C. A. (2002). Is face processing species-specific during the first year of life? *Science*, *296*(5571), 1321–1323. http://doi.org/10.1126/science.1070223

Paulesu, E., Vallar, G., Berlingeri, M., Signorini, M., Vitali, P., Burani, C., … Fazio, F. (2009). Supercalifragilisticexpialidocious: How the brain learns words never heard before. *NeuroImage*, *45*(4), 1368–1377. http://doi.org/10.1016/j.neuroimage.2008.12.043

Pearce, M. T., Müllensiefen, D., & Wiggins, G. A. (2010). The role of expectation and probabilistic learning in auditory boundary perception: A model comparison. *Perception*, *39*(10), 1365–1389. http://doi.org/10.1068/p6507

Peelle, J. E., Gross, J., & Davis, M. H. (2013). Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cerebral Cortex*, *23*(6), 1378–1387. http://doi.org/10.1093/cercor/bhs118

Peña, M., Bion, R. A. H., & Nespor, M. (2011). How modality specific is the iambic-trochaic law? Evidence from vision. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(5), 1199–1208. http://doi.org/10.1037/a0023944

Petersson, K. M., Folia, V., & Hagoort, P. (2012). What artificial grammar learning reveals about the neurobiology of syntax. *Brain and Language*, *120*(2), 83–95. http://doi.org/10.1016/j.bandl.2010.08.003

Petkov, C. I., Logothetis, N. K., & Obleser, J. (2009). Where are the human speech and voice regions, and do other animals have anything like them? *The Neuroscientist*, *15*(5), 419–429. http://doi.org/10.1177/1073858408326430

Petrov, A. A. (2011). Category rating is based on prototypes and not instances: Evidence from feedback-dependent context effects. *Journal of Experimental Psychology:*

Human Perception and Performance, *37*(2), 336–356. http://doi.org/10.1037/a0021436

Pickering, A. D. (1997). New approaches to the study of amnesic patients: What can a neurofunctional philosophy and neural network methods offer? *Memory*, *5*(1/2), 255–300. http://doi.org/10.1080/741941146

Pickett, E. R., Blumstein, S. E., & Burton, M. W. (1999). Effects of speaking rate on the singleton/geminate consonant contrast in Italian. *Phonetica*, *56*, 135–157.

Pickett, J. M., & Decker, L. R. (1960). Time factors in perception of a double consonant. *Language and Speech*, *3*(1), 11–17.

Pierrehumbert, J. B. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech*, *46*(2–3), 115–154. http://doi.org/10.1177/00238309030460020501

Pierrehumbert, J. B. (2016). Phonological representation: Beyond abstract versus episodic. *Annual Review of Linguistics*, *2*, 33–52. http://doi.org/10.1146/annurev-linguist-030514-125050

Pinard, M., Chertkow, H., Black, S., & Peretz, I. (2002). A case study of pure word deafness: Modularity in auditory processing? *Neurocase*, *8*(1–2), 40–55. http://doi.org/10.1093/neucas/8.1.40

Pind, J. (1995). Speaking rate, voice-onset time, and quantity: The search for higher-order invariants for two Icelandic speech cues. *Perception & Psychophysics*, *57*(3), 291–304. http://doi.org/10.3758/BF03213055

Pinel, P., Fauchereau, F., Moreno, A., Barbot, A., Lathrop, M., Zelenika, D., … Dehaene, S. (2012). Genetic variants of FOXP2 and KIAA0319/TTRAP/THEM2 locus are associated with altered brain activation in distinct language-related regions. *Journal of Neuroscience*, *32*(3), 817–825. http://doi.org/10.1523/JNEUROSCI.5996-10.2012

Pinker, S. (1998). Words and rules. *Lingua*, *106*(1), 219–242. http://doi.org/http://dx.doi.org/10.1016/S0024-3841(98)00035-7

Pinker, S. (1999). *Words and Rules: The Ingredients of Language*. Philadelphia, PA: Basic Books.

Pinker, S. (2005). So how does the mind work? *Mind and Language*, *20*(1), 1–24. http://doi.org/10.1111/j.0268-1064.2005.00274.x

Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition*, *28*(1–2), 73–193.

Pinker, S., & Ullman, M. T. (2002). The past and future of the past tense. *Trends in Cognitive Sciences*, *6*(11), 456–463. http://doi.org/10.1016/S1364-6613(02)01990-3

Pisoni, D. B., Aslin, R. N., Percy, A. J., & Hennessy, B. L. (1982). Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *Journal of Experimental Psychology: Human Perception and Performance*, *8*(2), 297–314.

Pitt, M. A., Szostak, C., & Dilley, L. C. (2016). Rate dependent speech processing can be speech specific: Evidence from the perceptual disappearance of words under changes in context speech rate. *Attention, Perception, & Psychophysics*, *78*, 334–345. http://doi.org/10.1177/0956797610384743

Pluymaekers, M., Ernestus, M., & Baayen, R. H. (2005). Lexical frequency and acoustic reduction in spoken Dutch. *Journal of the Acoustical Society of America*, *118*(4), 2561–2569. http://doi.org/10.1121/1.2011150

Poeppel, D. (2001). Pure word deafness and the bilateral processing of the speech code. *Cognitive Science*, *25*(5), 679–693. http://doi.org/10.1016/S0364-0213(01)00050-7

Poeppel, D., Idsardi, W. J., & van Wassenhove, V. (2008). Speech perception at the interface of neurobiology and linguistics. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *363*(1493), 1071–1086. http://doi.org/10.1098/rstb.2007.2160

Poldrack, R. A., & Foerde, K. (2008). Category learning and the memory systems debate. *Neuroscience and Biobehavioral Reviews*, *32*(2), 197–205. http://doi.org/10.1016/j.neubiorev.2007.07.007

Poole, J. H., Tyack, P. L., Stoeger-Horwath, A. S., & Watwood, S. (2005). Elephants are capable of vocal learning. *Nature*, *434*(7032), 455–456. http://doi.org/10.1029/2001GL014051

Port, R. F. (2007). How are words stored in memory? Beyond phones and phonemes. *New Ideas in Psychology*, *25*(2), 145–172. http://doi.org/10.1016/j.newideapsych.2007.02.001

Port, R. F., & Leary, A. P. (2005). Against formal phonology. *Language*, *81*(4), 927–964. http://doi.org/10.1.1.68.8395

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, *77*(3), 353–363. http://doi.org/10.1037/h0025953

Posner, M. I., & Keele, S. W. (1970). The retention of abstract ideas. *Journal of Experimental Psychology*, *83*(2), 304–308. http://doi.org/10.1037/h0028558

Price, C. J. (2012). A review and synthesis of the first 20years of PET and fMRI studies of heard speech, spoken language and reading. *NeuroImage*, *62*(2), 816–847.

http://doi.org/10.1016/j.neuroimage.2012.04.062

Price, C. J., Thierry, G., & Griffiths, T. (2005). Speech-specific auditory processing: Where is it? *Trends in Cognitive Sciences*, *9*(6), 271–276. http://doi.org/10.1016/j.tics.2005.03.009

Prince, A., & Smolensky, P. (2004). *Optimality Theory: Constraint interaction in generative grammar*. Malden, MA: Blackwell.

Proffitt, J. B., Coley, J. D., & Medin, D. L. (2000). Expertise and category-based induction. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(4), 811–828. http://doi.org/10.1037/0278-7393.26.4.811

Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nature Neuroscience*, *12*(6), 718–724. http://doi.org/10.1038/nn.2331

Reed, S. K. (1972). Pattern recognition and categorization. *Cognitive Psychology*, *3*(3), 382–407. http://doi.org/10.1016/0010-0285(72)90014-X

Reed, S. K., & Friedman, M. P. (1973). Perceptual vs conceptual categorization. *Memory & Cognition*, *1*(2), 157–163. http://doi.org/10.3758/BF03198087

Reinisch, E., Jesse, A., & McQueen, J. M. (2011). Speaking rate from proximal and distal contexts is used during word segmentation. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(3), 978–996. http://doi.org/10.1037/a0021923

Remez, R. E. (1989). When the objects of perception are spoken. *Ecological Psychology*, *1*(2), 161–180. http://doi.org/10.1207/s15326969eco0102

Remez, R. E., Pardo, J. S., Piorkowski, R. L., & Rubin, P. E. (2001). On the bistability of sine wave analogues of speech. *Psychological Science*, *12*(1), 24–29. http://doi.org/10.1111/1467-9280.00305

Remez, R. E., Rubin, P. E., Berns, S. M., Pardo, J. S., & Lang, J. M. (1994). On the perceptual organization of speech. *Psychological Review*, *101*(1), 129–156.

Remez, R. E., Rubin, P. E., Nygaard, L. C., & Howell, W. A. (1987). Perceptual normalization of vowels produced by sinusoidal voices. *Journal of Experimental Psychology: Human Perception and Performance*, *13*(1), 40–61.

Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. (1981). Speech perception without traditional speech cues. *Science*, *212*(4497), 947–950.

Repp, B. H. (1982). Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin*, *92*(1), 81–110. http://doi.org/10.1037/0033-2909.92.1.81

Repp, B. H., Milburn, C., & Ashkenas, J. (1983). Duplex perception: Confirmation of fusion. *Perception & Psychophysics*, *33*(4), 333–337. http://doi.org/10.3758/BF03205880

Reynolds, J. R., Zacks, J. M., & Braver, T. S. (2007). A computational model of event segmentation from perceptual prediction. *Cognitive Science*, *31*(4), 613–643. http://doi.org/10.1080/15326900701399913

Ripollés, P., Marco-Pallarés, J., Hielscher, U., Mestres-Missé, A., Tempelmann, C., Heinze, H. J., … Noesselt, T. (2014). The role of reward in word learning and its implications for language acquisition. *Current Biology*, *24*(21), 2606–2611. http://doi.org/10.1016/j.cub.2014.09.044

Rizzolatti, G., & Sinigaglia, C. (2010). The functional role of the parieto-frontal mirror circuit: interpretations and misinterpretations. *Nature Reviews: Neuroscience*, *11*, 264–274. http://doi.org/10.1038/nrn2805

Rocamora, M., López, E., & Jure, L. (2009). Wind instruments synthesis toolbox for generation of music audio signals with labeled partials. In *12th Brazilian Symposium on Computer Music*. Recife, Brazil.

Rosch, E. H. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, *104*(3), 192–233. http://doi.org/10.1037/0096-3445.104.3.192

Roseberry, S., Richie, R., Hirsh-Pasek, K., Golinkoff, R. M., & Shipley, T. F. (2011). Babies catch a break: 7- to 9-month-olds track statistical probabilities in continuous dynamic events. *Psychological Science*, *22*(11), 1422–1424. http://doi.org/10.1177/0956797611422074

Rosseel, Y. (2002). Mixture models of categorization. *Journal of Mathematical Psychology*, *46*(2), 178–210. http://doi.org/10.1006/jmps.2001.1379

Rumelhart, D. E., & Mcclelland, J. L. (1986). On learning the past tenses of English verbs. In J. L. McClelland & D. Rumelhart (Eds.), *Parallel Distributed Processing, Vol. 2* (pp. 535–551). Cambridge, Mass.

Saffran, E. M., Marin, O. S. M., & Yeni-Komshian, G. H. (1976). An analysis of speech perception in word deafness. *Brain and Language*, *3*(2), 209–228. http://doi.org/10.1016/0093-934X(76)90018-3

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928. http://doi.org/10.1126/science.274.5294.1926

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, *35*(4), 606–621. http://doi.org/10.1006/jmla.1996.0032

Samuel, A. G. (1982). Phonetic prototypes. *Perception & Psychophysics*, *31*(4), 307–314. http://doi.org/10.3758/BF03202653

Samuel, A. G., & Kraljic, T. (2009). Perceptual learning for speech. *Attention, Perception, & Psychophysics*, *71*(3), 1207–1218. http://doi.org/10.3758/APP

Sanborn, A. N., Griffiths, T. L., & Navarro, D. J. (2010). Rational approximations to rational models: Alternative algorithms for category learning. *Psychological Review*, *117*(4), 1144–1167. http://doi.org/10.1037/a0020511

Sargent, J. Q., Zacks, J. M., Hambrick, D. Z., Zacks, R. T., Kurby, C. A., Bailey, H. R., … Beck, T. M. (2013). Event segmentation ability uniquely predicts event memory. *Cognition*, *129*(2), 241–255. http://doi.org/10.1016/j.cognition.2013.07.002

Sawusch, J. R., & Newman, R. S. (2000). Perceptual normalization for speaking rate II: Effects of signal discontinuities. *Perception & Psychophysics, 62*(2), 285–300. http://doi.org/10.3758/BF03205549

Saygin, A. P., Leech, R., & Dick, F. (2010). Nonverbal auditory agnosia with lesion to Wernicke's area. *Neuropsychologia*, *48*(1), 107–113. http://doi.org/10.1016/j.neuropsychologia.2009.08.015

Saylor, M. M., Baldwin, D. A., Baird, J. A., & LaBounty, J. (2007). Infants' on-line segmentation of dynamic human action. *Journal of Cognition and Development*, *8*(1), 113–128. http://doi.org/10.1080/15248370709336996

Scharinger, M., Henry, M. J., & Obleser, J. (2013). Prior experience with negative spectral correlations promotes information integration during auditory category learning. *Memory & Cognition*, *41*(5), 752–768. http://doi.org/10.3758/s13421-013-0294-9

Scherf, K. S., Behrmann, M., Minshew, N., & Luna, B. (2008). Atypical development of face and greeble recognition in autism. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *49*(8), 838–847. http://doi.org/10.1111/j.1469-7610.2008.01903.x

Schmale, R., & Seidl, A. (2009). Accommodating variability in voice and foreign accent: Flexibility of early word representations. *Developmental Science*, *12*(4), 583–601. http://doi.org/10.1111/j.1467-7687.2009.00809.x

Schubotz, R. I., Korb, F. M., Schiffer, A. M., Stadler, W., & von Cramon, D. Y. (2012). The fraction of an action is more than a movement: Neural signatures of event segmentation in fMRI. *NeuroImage*, *61*(4), 1195–1205. http://doi.org/10.1016/j.neuroimage.2012.04.008

Schwartz, J.-L., Basirat, A., Ménard, L., & Sato, M. (2012). The Perception-for-Action-Control Theory (PACT): A perceptuo-motor theory of speech perception. *Journal of Neurolinguistics*, *25*(5), 336–354. http://doi.org/10.1016/j.jneuroling.2009.12.004

Scott, L. S., & Monesson, A. (2009). The origin of biases in face perception. *Psychological Science*, *20*(6), 676–680. http://doi.org/10.1111/j.1467-9280.2009.02348.x

Scott, S. K., Blank, C. C., Rosen, S., & Wise, R. J. S. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain*, *123*(12), 2400–2406. http://doi.org/10.1093/brain/123.12.2400

Scott, S. K., McGettigan, C., & Eisner, F. (2009). A little more conversation, a little less action—candidate roles for the motor cortex in speech perception. *Nature Reviews Neuroscience*, *10*(4), 295–302. http://doi.org/10.1038/nrn2603

Seger, C. A., & Miller, E. K. (2010). Category learning in the brain. *Annual Review of Neuroscience*, *33*, 203–219. http://doi.org/10.1146/annurev.neuro.051508.135546

Shafto, P., & Coley, J. D. (2003). Development of categorization and reasoning in the natural world: Novices to experts, naive similarity to ecological knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(4), 641–649. http://doi.org/10.1037/0278-7393.29.4.641

Shattuck-Hufnagel, S., & Turk, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research*, *25*(2), 193–247. http://doi.org/10.1007/BF01708572

Sheffert, S. M., Pisoni, D. B., Fellowes, J. M., & Remez, R. E. (2002). Learning to recognize talkers from natural, sinewave, and reversed speech samples. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(6), 1447–1469. http://doi.org/10.1037/0096-1523.28.6.1447

Shinn, P. C., Blumstein, S. E., & Jongman, A. (1985). Limitations of context conditioned effects in the perception of [b] and [w]. *Perception & Psychophysics*, *38*(5), 397–407. http://doi.org/10.3758/BF03207170

Shohamy, D., Myers, C. E., Kalanithi, J., & Gluck, M. A. (2008). Basal ganglia and dopamine contributions to probabilistic category learning. *Neuroscience and Biobehavioral Reviews*, *32*(2), 219–236. http://doi.org/10.1016/j.neubiorev.2007.07.008

Shtyrov, Y. (2012). Neural bases of rapid word learning. *The Neuroscientist*, *18*(4), 312–319. http://doi.org/10.1177/1073858411420299

Shtyrov, Y., Nikulin, V. V, & Pulvermuller, F. (2010). Rapid cortical plasticity underlying novel word learning. *Journal of Neuroscience*, *30*(50), 16864–16867. http://doi.org/10.1523/JNEUROSCI.1376-10.2010\r30/50/16864 [pii]

Sidaras, S. K., Alexander, J. E. D., & Nygaard, L. C. (2009). Perceptual learning of systematic variation in Spanish-accented speech. *Journal of the Acoustical Society of America*, *125*(5), 3306–3316. http://doi.org/10.1121/1.3101452

Sinex, D. G., McDonald, L. P., & Mott, J. B. (1991). Neural correlates of nonmonotonic temporal acuity for voice onset time. *Journal of the Acoustical Society of America*, *90*(5), 2441–2449. http://doi.org/10.1121/1.402048

Sinnott, J. M., & Brown, C. H. (1997). Perception of the American English liquid /ra-la/ contrast by humans and monkeys. *Journal of the Acoustical Society of America*, *102*(1), 588–602. http://doi.org/10.1121/1.419732

Sinnott, J. M., Brown, C. H., Malik, W. T., & Kressley, R. A. (1997). A multidimensional scaling analysis of vowel discrimination in humans and monkeys. *Perception & Psychophysics*, *59*(8), 1214–1224. http://doi.org/10.3758/BF03214209

Sinnott, J. M., & Saporita, T. A. (2000). Differences in American English, Spanish, and monkey perception of the say-stay trading relation. *Perception & Psychophysics*, *62*(6), 1312–1319. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11019626

Sinnott, J. M., & Williamson, T. L. (1999). Can macaques perceive place of articulation from formant transition information? *Journal of the Acoustical Society of America*, *106*(2), 929–937. http://doi.org/10.1121/1.427107

Slote, J., & Strand, J. F. (2016). Conducting spoken word recognition research online: Validation and a new timing method. *Behavior Research Methods*, *48*(2), 553–566. http://doi.org/10.3758/s13428-015-0599-7

Smayda, K. E., Chandrasekaran, B., & Maddox, W. T. (2015). Enhanced cognitive and perceptual processing: A computational basis for the musician advantage in speech learning. *Frontiers in Psychology*, *6*, 682. http://doi.org/10.3389/fpsyg.2015.00682

Smith, E. E., & Grossman, M. (2008). Multiple systems of category learning. *Neuroscience and Biobehavioral Reviews*, *32*(2), 249–264. http://doi.org/10.1016/j.neubiorev.2007.07.009

Smith, E. E., Patalano, A. L., & Jonides, J. (1998). Alternative strategies of categorization. *Cognition*, *65*(2–3), 167–196. http://doi.org/10.1016/S0010-0277(97)00043-7

Smith, J. D., & Minda, J. P. (1998). Prototypes in the mist: The early epochs of category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(6), 1411–1436. http://doi.org/10.1037//0278-7393.24.6.1411

Soha, J. A., & Peters, S. (2015). Vocal learning in songbirds and humans: A retrospective in honor of Peter Marler. *Ethology*, *121*(10), 933–945. http://doi.org/10.1111/eth.12415

Speer, N. K., & Zacks, J. M. (2005). Temporal changes as event boundaries: Processing and memory consequences of narrative time shifts. *Journal of Memory and Language*, *53*, 125–140. http://doi.org/10.1016/j.jml.2005.02.009

Sperber, D. (2001). In defense of massive modularity. In E. Dupoux (Ed.), *Language, Brain, and Cognitive Development: Essays in Honor of Jacques Mehler* (pp. 47–57).

Cambridge, Mass.: MIT Press.

Spreen, O., Benton, A. L., & Fincham, R. W. (1965). Auditory agnosia without aphasia. *Archives of Neurology*, *13*(1), 84–92.

Squire, L. R. (2009). Memory and brain systems: 1969-2009. *Journal of Neuroscience*, *29*(41), 12711–12716. http://doi.org/10.1523/JNEUROSCI.3575-09.2009

Sridharan, D., Levitin, D. J., Chafe, C. H., Berger, J., & Menon, V. (2007). Neural dynamics of evetn segmentation in music: Converging evidence for dissociable ventral and dorsal networks. *Neuron*, *55*(3), 521–532. http://doi.org/10.1016/j.neuron.2007.07.003

Stahl, A. E., Romberg, A. R., Roseberry, S., Golinkoff, R. M., & Hirsh-Pasek, K. (2014). Infants segment continuous events using transitional probabilities. *Child Development*, *85*(5), 1821–1826. http://doi.org/10.1111/cdev.12247

Stains, M., & Talanquer, V. (2008). Classification of chemical reactions: Stages of expertise. *Journal of Research in Science Teaching*, *45*(7), 771–793. http://doi.org/10.1002/tea.20221

Steinschneider, M., Nourski, K. V, & Fishman, Y. I. (2013). Representation of speech in human auditory cortex: Is it special? *Hearing Research*, *305*(1), 57–73. http://doi.org/10.1016/j.heares.2013.05.013

Stevens, K. N., & Blumstein, S. E. (1978). Invariant cues for place of articulation in stop consonants. *Journal of the Acoustical Society of America*, *64*(5), 1358–1368. http://doi.org/10.1121/1.382102

Stickgold, R. (2005). Sleep-dependent memory consolidation. *Nature*, *437*(7063), 1272–1278. http://doi.org/10.1038/nature04286

Storkel, H. L. (2015). Learning from input and memory evolution: Points of vulnerability on a pathway to mastery in word learning. *International Journal of Speech-Language Pathology*, *17*(1), 1–12. http://doi.org/10.3109/17549507.2014.987818

Storms, G., De Boeck, P., & Ruts, W. (2001). Categorization of novel stimuli in well-known natural concepts: A case study. *Psychonomic Bulletin & Review*, *8*(2), 377–384. http://doi.org/10.3758/BF03196176

Summerfield, Q. (1981). Articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance*, *7*(5), 1074–1095. http://doi.org/10.1037/0096-1523.7.5.1074

Sussman, H. M., 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*(3), 1309–1325. http://doi.org/10.1121/1.401923

Swallow, K. M., Zacks, J. M., & Abrams, R. A. (2009). Event boundaries in perception affect memory encoding and updating. *Journal of Experimental Psychology: General*, *138*(2), 236–257. http://doi.org/10.1037/a0015631

Tamminen, J., Davis, M. H., Merkx, M., & Rastle, K. (2012). The role of memory consolidation in generalisation of new linguistic information. *Cognition*, *125*(1), 107–112. http://doi.org/10.1016/j.cognition.2012.06.014

Tanaka, Y., Yamadori, A., & Mori, E. (1987). Pure word deafness following bilateral lesions: A psychophysical analysis. *Brain*, *110*(2), 381–403. http://doi.org/10.1093/brain/110.2.381

Taniwaki, T., Tagawa, K., Sato, F., & Iino, K. (2000). Auditory agnosia restricted to environmental sounds following cortical deafness and generalized auditory agnosia. *Clinical Neurology and Neurosurgery*, *102*(3), 156–162. http://doi.org/10.1016/S0303-8467(00)00090-1

Tarr, M., & Gauthier, I. (2000). FFA: a flexible fusiform area for subordinate-level visual processing automatized by expertise. *Nature Neuroscience*, *3*(8), 764–769. http://doi.org/10.1038/77666

Tauzin, T. (2015). Simple visual cues of event boundaries. *Acta Psychologica*, *158*, 8–18. http://doi.org/10.1016/j.actpsy.2015.03.007

Tenenbaum, J. B., Griffiths, T. L., & Kemp, C. (2006). Theory-based Bayesian models of inductive learning and reasoning. *Trends in Cognitive Sciences*, *10*(7), 309–318. http://doi.org/10.1016/j.tics.2006.05.009

Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy*, *7*(1), 53–71. http://doi.org/10.1207/s15327078in0701_5

Thiessen, E. D., & Saffran, J. R. (2003). When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Developmental Psychology*, *39*(4), 706–716. http://doi.org/10.1037/0012-1649.39.4.706

Tooby, J., & Cosmides, L. (1990). On the universality of human nature and the uniqueness of the individual: The role of genetics and adaptation. *Journal of Personality*, *58*(1), 17–67. http://doi.org/10.1111/j.1467-6494.1990.tb00907.x

Toscano, J. C., & McMurray, B. (2010). Cue integration with categories: Weighting acoustic cues in speech using unsupervised learning and distributional statistics. *Cognitive Science*, *34*(3), 434–464. http://doi.org/10.1111/j.1551-

6709.2009.01077.x.Cue

Trope, Y., & Liberman, N. (2003). Temporal construal. *Psychological Review*, *110*(3), 403–421. http://doi.org/10.1037/0033-295X.110.3.403

Trout, J. D. (2001). The biological basis of speech: what to infer from talking to the animals. *Psychological Review*, *108*(3), 523–549. http://doi.org/10.1037/0033-295X.108.3.523

Turk, A. E., & Shattuck-Hufnagel, S. (2000). Word-boundary-related duration patterns in English. *Journal of Phonetics*, *28*(4), 397–440. http://doi.org/10.006/jpho.2000.0123

Turk, A. E., & Shattuck-Hufnagel, S. (2014). Timing in talking: What is it used for, and how is it controlled? *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *369*(1658), 1–13. http://doi.org/10.1098/rstb.2013.0395

Tversky, B., & Zacks, J. M. (2013). Event perception. In D. Reisberg (Ed.), *Oxford Handbook of Cognitive Psychology* (pp. 83–95). Oxford: Oxford University Press. http://doi.org/10.1002/wcs.133

Ullman, M. T. (1999). Acceptability ratings of regular and irregular past-tense forms: Evidence for a dual-system model of language from word frequency and phonological neighbourhood effects. *Language and Cognitive Processes*, *14*(1), 47–67. http://doi.org/10.1080/016909699386374

Ullman, M. T. (2001). The neural basis of lexicon and grammar in first and second language: The declarative/procedural model. *Bilingualism: Language and Cognition*, *4*(1), 105–122. http://doi.org/10.1017/S1366728901000220

Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition*, *92*(1–2), 231–270.

http://doi.org/10.1016/j.cognition.2003.10.008

Ullman, M. T. (2016). The declarative/procedural model: A neurobiological model of language learning, knowledge, and use. In G. Hickok & S. L. Small (Eds.), *Neurobiology of Language* (pp. 953–968). Amsterdam: Elsevier.

Ullman, M. T., Corkin, S., Coppola, M., Hickok, G., Growdon, J. H., Koroshetz, W. J., & Pinker, S. (1997). A neural dissociation within language: Evidence that the mental dictionary is part of declarative memory, and that grammatical rules are processed by the procedural system. *Journal of Cognitive Neuroscience*, *9*(2), 266–276. http://doi.org/10.1162/jocn.1997.9.2.266

Ullman, M. T., & Pullman, M. Y. (2015). A compensatory role for declarative memory in neurodevelopmental disorders. *Neuroscience and Biobehavioral Reviews*, *51*, 205–222. http://doi.org/10.1016/j.neubiorev.2015.01.008

van Heugten, M., & Johnson, E. K. (2014). Learning to contend with accents in infancy: Benefits of brief speaker exposure. *Journal of Experimental Psychology: General*, *143*(1), 340–350. http://doi.org/10.1037/a0032192

Vandierendonck, A. (1995). A parallel rule activation and rule synthesis model for generalization in category learning. *Psychonomic Bulletin & Review*, *2*(4), 442–459. http://doi.org/10.3758/BF03210982

Vargha-Khadem, F., Gadian, D. G., Watkins, K. E., Connelly, A., Van Paesschen, W., & Mishkin, M. (1997). Differential effects of early hippocampal pathology on episodic and semantic memory. *Science*, *277*(5324), 376–380. http://doi.org/10.1126/science.277.5324.376

Verbrugge, R. R., Strange, W., Shankweiler, D. P., & Edman, T. R. (1976). What information enables a listener to map a talker's vowel space? *Journal of the Acoustical Society of America*, *60*(1), 198–212. http://doi.org/10.1121/1.381065

Vermaercke, B., Cop, E., Willems, S., D'Hooge, R., & Op de Beeck, H. P. (2014). More complex brains are not always better: Rats outperform humans in implicit category-based generalization by implementing a similarity-based strategy. *Psychonomic Bulletin & Review*, *21*(4), 1080–1086. http://doi.org/10.3758/s13423-013-0579-9

Viswanathan, N., Magnuson, J. S., & Fowler, C. A. (2010). Compensation for coarticulation: disentangling auditory and gestural theories of perception of coarticulatory effects in speech. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(4), 1005–1015. http://doi.org/10.1037/a0018391

Volaitis, L. E., & Miller, J. L. (1992). Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories. *Journal of the Acoustical Society of America*, *92*(2), 723–735. http://doi.org/10.1121/1.403997

Vouloumanos, A., & Curtin, S. (2014). Foundational tuning: How infants' attention to speech predicts language development. *Cognitive Science*, *38*(8), 1675–1686. http://doi.org/10.1111/cogs.12128

Vouloumanos, A., Martin, A., & Onishi, K. H. (2014). Do 6-month-olds understand that speech can communicate? *Developmental Science*, *17*(6), 872–879. http://doi.org/10.1111/desc.12170

Vouloumanos, A., & Werker, J. F. (2004). Tuned to the signal: The privileged status of speech for young infants. *Developmental Science*, *7*(3), 270–276. http://doi.org/10.1111/j.1467-7687.2004.00345.x

Vouloumanos, A., & Werker, J. F. (2007). Listening to language at birth: Evidence for a bias for speech in neonates. *Developmental Science*, *10*(2), 159–164. http://doi.org/10.1111/j.1467-7687.2007.00549.x

Wade, T., & Holt, L. L. (2005). Perceptual effects of preceding nonspeech rate on temporal properties of speech categories. *Perception & Psychophysics*, *67*(6), 939–950. http://doi.org/10.3758/BF03193621

Waldron, E., & Ashby, F. G. (2001). The effects of concurrent task interference on category learning: evidence for multiple category learning systems. *Psychonomic Bulletin & Review*, *8*(1), 168–176.

Wang, H.-C., Savage, G., Gaskell, M. G., Paulin, T., Robidoux, S., & Castles, A. (2016). Bedding down new words: Sleep promotes the emergence of lexical competition in visual word recognition. *Psychonomic Bulletin & Review*. http://doi.org/10.3758/s13423-016-1182-7

Watkins, K. E., Smith, S. M., Davis, S., & Howell, P. (2008). Structural and functional abnormalities of the motor system in developmental stuttering. *Brain*, *131*(1), 50–59. http://doi.org/10.1093/brain/awm241

Wayland, S. C., Miller, J. L., & Volaitis, L. E. (1994). The influence of sentential speaking rate on the internal structure of phonetic categories categories. *Journal of the Acoustical Society of America*, *95*(5), 2694–2701. http://doi.org/10.1121/1.409838

Werker, J. F., & Lalonde, C. E. (1988). Cross-language speech perception: Initial capabilities and developmental change. *Developmental Psychology*, *24*(5), 672–683. http://doi.org/10.1037/0012-1649.24.5.672

Westbury, J. R., Hashi, M., & Lindstrom, M. J. (1998). Differences among speakers in lingual articulation for American English /r/. *Speech Communication*, *26*(3), 203–226. http://doi.org/10.1016/S0167-6393(98)00058-2

Whalen, D. H., & Liberman, A. M. (1987). Speech perception takes precedence over nonspeech perception. *Science*, *237*(4811), 169–171.

White, K. S., & Aslin, R. N. (2011). Adaptation to novel accents by toddlers. *Developmental Science*, *14*(2), 372–384. http://doi.org/10.1111/j.1467-7687.2010.00986.x

Wickelgren, W. A. (1965). Distinctive features and errors in short-term memory for English vowels. *Journal of the Acoustical Society of America*, *38*(4), 583–588.

Wickelgren, W. A. (1966). Distinctive features and errors in short-term memory for English consonants. *Journal of the Acoustical Society of America*, *39*(2), 388–398.

Wieland, E. A., McAuley, J. D., Dilley, L. C., & Chang, S. E. (2015). Evidence for a rhythm perception deficit in children who stutter. *Brain and Language*, *144*, 26–34. http://doi.org/10.1016/j.bandl.2015.03.008

Willems, R. M., & Hagoort, P. (2007). Neural evidence for the interplay between language, gesture, and action: A review. *Brain and Language*, *101*(3), 278–289. http://doi.org/10.1016/j.bandl.2007.03.004

Worgan, S. F., & Moore, R. K. (2010). Speech as the perception of affordances. *Ecological Psychology*, *22*(4), 327–343. http://doi.org/10.1080/10407413.2010.517125

Yang, J., & Li, P. (2012). Brain networks of explicit and implicit learning. *PLoS ONE*, *7*(8), e42993. http://doi.org/10.1371/journal.pone.0042993

Yi, H.-G., Maddox, W. T., Mumford, J. A., & Chandrasekaran, B. (2016). The role of corticostriatal systems in speech category learning. *Cerebral Cortex*, *26*(4), 1409–1420. http://doi.org/10.1093/cercor/bhu236

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology, 81*(1), 141–145. http://doi.org/10.1037/h0027474

Yovel, G., & Kanwisher, N. G. (2004). Face perception: Domain specific, not process specific. *Neuron*, *44*(5), 889–898. http://doi.org/10.1016/S0896-6273(04)00728-7

Zacks, J. M. (2004). Using movement and intentions to understand simple events. *Cognitive Science*, *28*(6), 979–1008. http://doi.org/10.1016/j.cogsci.2004.06.003

Zacks, J. M., Kumar, S., Abrams, R. A., & Mehta, R. (2009). Using movement and intentions to understand human activity. *Cognition*, *112*(2), 201–216. http://doi.org/10.1016/j.cognition.2009.03.007

Zacks, J. M., Kurby, C. A., Landazabal, C. S., Krueger, F., & Grafman, J. (2016). Effects of penetrating traumatic brain injury on event segmentation and memory. *Cortex, 74*, 233–246. http://doi.org/10.1016/j.cortex.2015.11.002

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: a mind-brain perspective. *Psychological Bulletin*, *133*(2), 273–293. http://doi.org/10.1037/0033-2909.133.2.273

Zacks, J. M., & Swallow, K. M. (2007). Event segmentation. *Current Directions in Psychological Science*, *16*(2), 80–84. http://doi.org/10.1111/j.1467-8721.2007.00480.x

Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin*, *127*(1), 3–21. http://doi.org/10.1037//0033-2909.127.1.3

Zeithamova, D., & Maddox, W. T. (2006). Dual-task interference in perceptual category learning. *Memory & Cognition*, *34*(2), 387–398.